Vol. 30, No. 2 April, 1946

Journal of Applied Psychology

EDITED BY: DONALD G. PATERSON, UNIVERSITY OF MINNESOTA

Consulting Editors

PAUL S. ACHILLES, *Psychological Corporation;* WALTER V. BINGHAM, *A.G.O., War Department;* HAROLD E. BURTT, *Ohio State University;* ARTHUR I. GATES, *T. C. Columbia University;* JOHN G. JENKINS, *University of Maryland;* IRVING LORGE, *T. C. Columbia University;* QUINN MCNEMAR, *Stanford University;* WILLARD C. OLSON, *University of Michigan;* JAMES P. PORTER, *Swarthmore, Pennsylvania;* EDWARD K. STRONG, JR., *Stanford University;* MORRIS S. VITELES, *University of Pennsylvania;* JOSEPH ZUBIN, *N. Y. Psychiatric Institute.*

Table of Contents

Studies in Job Evaluation. 3. An Analysis of Point Ratings for Salary Paid Jobs in an Industrial Plant: C. H. LAWSHE, JR., AND A. A. MALESKI 117

A Preliminary Report on the Miami-Oxford Curve-Block Series: P. L. MELLENBRUCH 129

A Comparative Study of Three Tests for Color Vision: H. FOSTER 135

Studies in the Application of Motor Skills Techniques to the Vocational Adjustment of the Blind: M. K. BAUMAN 144

The Significance of Verbal Aptitude in the Type of Occupation Pursued by Illiterates: W. D. ALTUS AND C. A. MAHLER 155

Readability of Newspaper Headlines Printed in Capitals and in Lower Case: D. G. PATERSON AND M. A. TINKER 161

Explorations in Personality by the Sentence Completion Method: A. R. ROHDE 169

Speech Intelligibility Under Various Degrees of Anoxia: G. M. SMITH AND C. P. SEITZ 182

Book Reviews 192

New Books, Monographs, and Pamphlets 196

Published Bi-monthly by The American Psychological Association, Inc.

Prince and Lemon Sts., Lancaster, Pa., and 1227 Nineteenth St., NW, Washington 6, D. C.

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the Act of March 3, 1879

Copyright, 1946, by The American Psychological Association, Inc.

Journal of Applied Psychology

Vol. 30, No. 2 April, 1946

Studies in Job Evaluation. 3. An Analysis of Point Ratings for Salary Paid Jobs in an Industrial Plant

C. H. Lawshe, Jr., and A. A. Maleski

Division of Education and Applied Psychology, Purdue University

The growing importance of job evaluation, the various approaches to the problem, and the increasing concern of the applied psychologist with this area of industrial activity have been discussed in earlier studies [1] in this series. These studies have dealt with point rating systems as they operate with hourly paid jobs, while the present one deals with a modified system used in the evaluation of salary paid jobs. The findings and interpretations in this study apply specifically to a salary rating plan in a particular plant; however, certain general principles can be found which may be inferred in other industiral situations.

Purpose of the Study

The principal purpose of this study is two-fold: (1) to identify through statistical techniques the primary factors operating in a salary rating plan as it is functioning in an industrial plant, to discover those items which tend to gravitate or cluster around the factors defined, and to determine the significance of each factor in the total point rating; (2) to determine the adequacy of an abbreviated point rating scale based on a few items, and to disclose and study any differences between the two scales in terms of their practical significance.

Rating Scale and Experimental Procedure

Salary Rating Plan. To enable its member companies to evaluate salaried positions in the same objective way that factory jobs can be rated, the National Metal Trades Association devised a salary rating

[1] C. H. Lawshe, Jr., and G. A. Satter. Studies in job evaluation. 1. Factor analyses of point ratings for hourly-paid jobs in three industrial plants. *Journal of Applied Psychology*, Vol. 28, No. 3, pp. 189–198, June, 1944. C. H. Lawshe, Jr. Studies in job evaluation. 2. The adequacy of abbreviated point ratings for hourly paid jobs in three industrial plants. *Journal of Applied Psychology*, Vol. 29, No. 3, pp. 177–184, 1945.

plan which is similar in its structure and method to the N. E. M. A. system of point rating as reported by Kress.[2] This plan provides for the rating of salary paid jobs on the following items: Education; Experience; Complexity of Duties; Supervision Received; Working Conditions; Errors; Contacts with Others; Confidential Data; and Mental or Visual Demand. The plan also includes two items which are used only when rating positions where supervisory duties are involved. These are "Scope of Supervision" and "Character of Supervision." This means that all other jobs receive a "zero" on the two items; actually, all jobs are rated on these items just as they are on the others.

Each item is graded into 5, 6, or 7 degrees and each degree carries a different numerical or point value. For example, the item "Scope of Supervision" appraises the magnitude of the supervisory responsibility expressed in terms of the number of persons supervised and is graded into six categories as follows:

Degree	*Description*	*Points*
1	Assist and direct up to 5 persons usually in the same occupation.	5
2	Supervise a small group or unit, seldom over 25 persons.	10
3	Supervise a section or department, seldom over 50 persons.	20
4	Supervise a department, from 50 to seldom over 100 persons.	40
5	Direct and supervise one or more departments, usually from 100 to 250 persons.	60
6	Direct and supervise the operations of a major division, usually more than 250 persons.	80

Points awarded to a salaried position on the item, "Scope of Supervision," are added to the values derived from the other ten items on the same job to give a total point rating for that particular job. Thus by rating each position in a plant or organization on the eleven item scale, the relative worth of each job can be expressed quantitatively in terms of total points. These totals are then converted into money values to establish a salary structure for the organization.

Source of Data. Point rating data were obtained on about 400 different salaried jobs from an industrial plant manufacturing airframes and employing over 5,000 workers. This plant also has a point rating plan in operation for its hourly-paid employees, which has been studied by Lawshe and Satter[3] as Plant B.

Procedure. The salary rating data including the ratings on each of the eleven items plus the total point ratings for each job were punched on

[2] A. L. Kress. How to rate jobs and men. *Factory Management*, 60–65, 1939.
[3] C. H. Lawshe, Jr., and G. A. Satter. *Op. cit.*

Table 1

Intercorrelations of Point Ratings of Eleven Items and of Total Points in the Job Evaluation of Salaried Workers in an Industrial Plant

	(1) Total Points	(2) Education	(3) Experience	(4) Complexity of Duties	(5) Supervision Received	(6) Errors	(7) Contacts with Others	(8) Confidential Data	(9) Mental or Visual Demand	(10) Working Conditions	(11) Character of Supervision
(2) Education	.809										
(3) Experience	.930	.756									
(4) Complexity of Duties	.928	.788	.856								
(5) Supervision Received	.860	.642	.762	.783							
(6) Errors	.864	.677	.772	.788	.771						
(7) Contacts with Others	.856	.657	.750	.768	.754	.761					
(8) Confidential Data	.717	.617	.667	.663	.607	.591	.642				
(9) Mental or Visual Demand	−.346	−.174	−.259	−.335	−.364	−.337	−.419	−.149			
(10) Working Conditions	.113	−.144	.071	.063	.097	.096	.108	−.095	−.290		
(11) Character of Supervision	.789	.494	.658	.667	.690	.630	.641	.496	−.350	.195	
(12) Scope of Supervision	.597	.289	.465	.489	.485	.429	.410	.280	−.291	.253	.708

Table 2

Factor Loadings Before and After Rotation

Rating Scale Items	Before Rotation				After Rotation			
	k_1	k_2	k_3	h^2	k_1	k_2	k_3	h^2
(1) Total Points	.991	.126	−.095	1.007	.982	.194	.081	1.009
(2) Education	.745	.436	−.127	.761	.844	−.022	−.223	.763
(3) Experience	.893	.208	−.093	.849	.914	.122	.000	.850
(4) Complexity of Duties	.914	.178	−.056	.870	.924	.117	.051	.870
(5) Supervision Received	.870	.045	.080	.765	.841	.086	.225	.765
(6) Errors	.860	.135	.149	.780	.860	−.024	.202	.781
(7) Contacts with Others	.864	.098	.213	.802	.852	−.049	.271	.802
(8) Confidential Data	.686	.330	−.069	.584	.755	−.013	−.122	.585
(9) Mental or Visual Demand	−.423	.322	−.318	.384	−.302	−.041	−.539	.383
(10) Working Conditions	.152	−.467	.178	.273	.000	.183	.489	.273
(11) Character of Supervision	.806	−.305	−.240	.800	.672	.533	.256	.801
(12) Scope of Supervision	.613	−.439	−.298	.657	.447	.621	.270	.658

machine-sort cards. Intercorrelations between all eleven items and total points were computed and a correlation matrix was prepared (Table 1). The matrix was factor analyzed by Thurstone's centroid method and three factors (k_1, k_2 and k_3) were extracted. The extraction process was stopped when the communalities (h^2) for "total points" approximated unity (1.007). The factors and their centroid loadings are presented in Table 2 and the factor names with item groupings in Table 3. The factor loadings were transformed by the method outlined by Peters and Van Voorhis,[4] and transformed or rotated values are also found in Table 2.

The Wherry-Doolittle shrinkage selection method was applied to the values in the correlation matrix and the three "best" items were selected for an abbreviated rating scale.

Identification of Factors

Factor Names. Of the three primary factors identified through factor analysis, Factor I was found to be common in some degree to all of the items except "working conditions." However, eight of the items stand out as possessing particularly heavy loadings in Factor I (see Table 2). These eight items are presented in Table 3 in the order of the magnitude of their loadings which range from .924 for "complexity of duties" to .672 for "character of supervision." The items identified

Table 3

Three Factors with Rating Scale Items Arranged in Order of Magnitude of Loadings

Factor	Rating Scale Item	Loading
I. Skill Demands	(4) Complexity of Duties	.924
	(3) Experience	.914
	(6) Errors	.860
	(7) Contacts with Others	.852
	(2) Education	.844
	(5) Supervision Received	.841
	(8) Confidential Data	.755
	(11) Character of Supervision*	.672
II. Supervisory Demands	(12) Scope of Supervision	.621
	(11) Character of Supervision*	.533
III. Job Characteristics	(9) Mental or Visual Demand	−.539
	(10) Working Conditions	.489

* The item Character of Supervision (11) has a relatively heavy loading in both Factors I and II.

[4] C. C. Peters and W. R. Van Voorhis. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill Book Co., 264–268, 1940.

under Factor I all embrace some skill requirement of the individual who is to perform the job successfully and therefore it has been named "Skill Demands."

The items "scope of supervision" and "character of supervision" are heavily weighted with Factor II (see Table 2), their respective loadings being .621 and .533. These items were found to function in jobs where supervisory duties are involved. They also represent a skill demand, but it is obviously a more specific and specialized type of skill requiring supervisory ability, so Factor II has been named "Supervisory Demands." The item "character of supervision" is significantly loaded with Factors I and II and is listed under both in Table 3, since the extent of the contribution of each seems to be approximately the same.

The largest loadings of Factor III are − .539 in the item "mental or visual demand" and .489 in the item "working conditions." This factor was named "Job Characteristics" since the two items represent certain aspects of the job with which an individual must contend, and for which he should be compensated. It is significant that the item "mental or visual demand" is negatively loaded with Factor III (see Tables 2 and 3). This negative loading although not too common in factor analysis, is quite plausible and is discussed by Peters and Van Voorhis.[5] As shown in Table 1 "mental or visual demand" correlates negatively not only with "total points" but also with each of the other ten items. In other words, whenever a salaried position in this plant was assigned a higher point value on the item "mental or visual demand," it generally received a relatively lower point value on the other items and consequently on "total points." Therefore, this item plays an important role in Factor III but in a negative direction. It is possible, by reversing the signs for "mental or visual demand," to get a positive loading. However, this would not affect the absolute value in any way, so perhaps from the logical point of view it is more meaningful to allow this loading to remain negative.

Factor Significance. Table 2 shows that the communality (h_2) for "total points" is 1.009, which since it approximates unity, indicates that the three factors extracted account for just about all of the variability in the total point ratings. If it can be assumed that these are the best rotations, then the relative proportions that each factor contributes to the total variability may be computed from the squares of the total point loadings. Thus Factor I, which has a total point loading of .982, accounts for most of the variability. Figure 1 presents these proportions graphically. Factor I, "Skill Demands" accounts for 95.6% of the total variability. Factor II, "Supervisory Demands" 3.7% and Factor III, "Job Characteristics," only .7%.

[5] C. C. Peters and W. R. Van Voorhis. *Ibid.*, pp. 270–272.

The results in the identification of the basic factors in this scale are essentially similar to those reported by Lawshe and Satter [6] in their study of point ratings on hourly-paid jobs in three different industrial plants. They found the factors "Skill Demands" and "Job Characteristics" present in all three plants and that "Skill Demands" accounted for most of the variance in the total point ratings in each plant. The

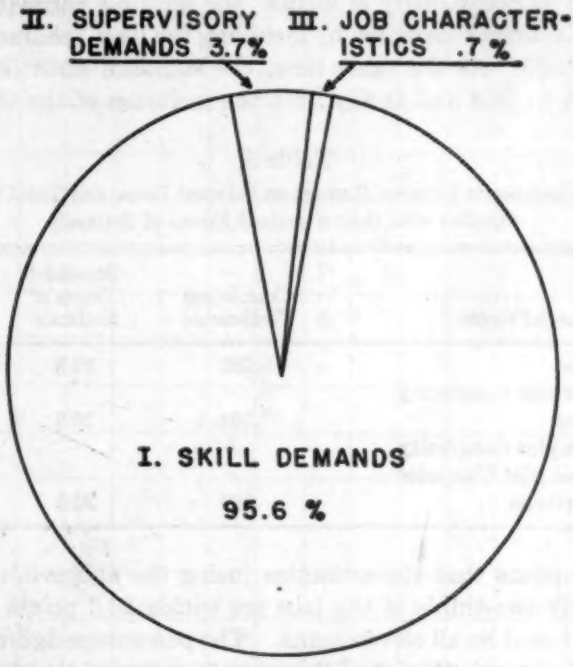


FIG. 1. The relative proportions for each of the factor contributions to the total point ratings of salary positions.

factor "Supervisory Demands" did not appear in the study. However, since it is presumably characteristic primarily of salary paid positions it would not be expected to appear in hourly-paid jobs.

Abbreviated Rating Scale

Items Selected. Of the original eleven items, the three items selected by the Wherry-Doolittle technique for an abbreviated rating scale were "experience," "complexity of duties," and "character of supervision."

They are presented with their multiple *R*'s and standard errors of estimate in Table 4. The selection process was terminated with these three items since the trial addition of a fourth variable increased the

[6] C. H. Lawshe, Jr., and G. A. Satter. *Op. cit.*

multiple R only from .980 to .987 or less than .01. This increase does not appear significant either from a statistical point of view or in terms of sheer practicality.

Accuracy of Prediction. It can be seen from Table 4 that the variable "experience" correlates highest with "total points" having a coefficient of .930 and a standard error of estimate of 32.3 points. However, with the addition of "complexity of duties" the multiple correlation is raised to .964 and is further increased by including the item "character of supervision," to .980. At the same time, the standard error of estimate is reduced first to 26.8 and finally with the inclusion of the third item, to

Table 4

Correlation Coefficients Between Ratings on Selected Items and Total Point Ratings together with their Standard Errors of Estimate

Selected Items	Correlation Coefficients	Standard Errors of Estimate	%
Experience	.930	32.3	.32
Experience plus Complexity of Duties	.964	26.8	.26
Experience plus Complexity of Duties plus Character of Supervision	.980	20.3	.20

20.3. This means that the estimates (using the abbreviated scale) for approximately two-thirds of the jobs are within 20.3 points of the total point rating based on all eleven items. The percentage figures in Table 4 indicate the proportional size of the errors in terms of the standard deviation of the total point distribution.

Relative Contribution to Total Point Ratings. The square of the multiple $R(.980^2)$ for the three item scale is .96 which indicates that about 96% of the variance in "total points" can be attributed to the combined effect of "experience," "complexity of duties," and "character of supervision." The other eight items apparently contribute a total of 4%.

Application of the Abbreviated Scale. To determine the extent to which the abbreviated scale would give results the same as or comparable to the original scale, an analysis was made of the changes that would occur if it were actually used to rate the same positions. This was done by developing a regression equation for predicting "total points" from the three selected items as follows:

$$X_{TP} = 15 + 1.8_{\text{Exp.}} + 2.7_{CD} + 2.1_{CS}.$$

Point ratings on "experience," "complexity of duties," and "character of supervision" were substituted in the formula for each of the salaried jobs and the constant 15 added to obtain the estimated or computed total point ratings.

The computed total point values were plotted against the original total point ratings in the scattergram in Figure 2. In the salary plan

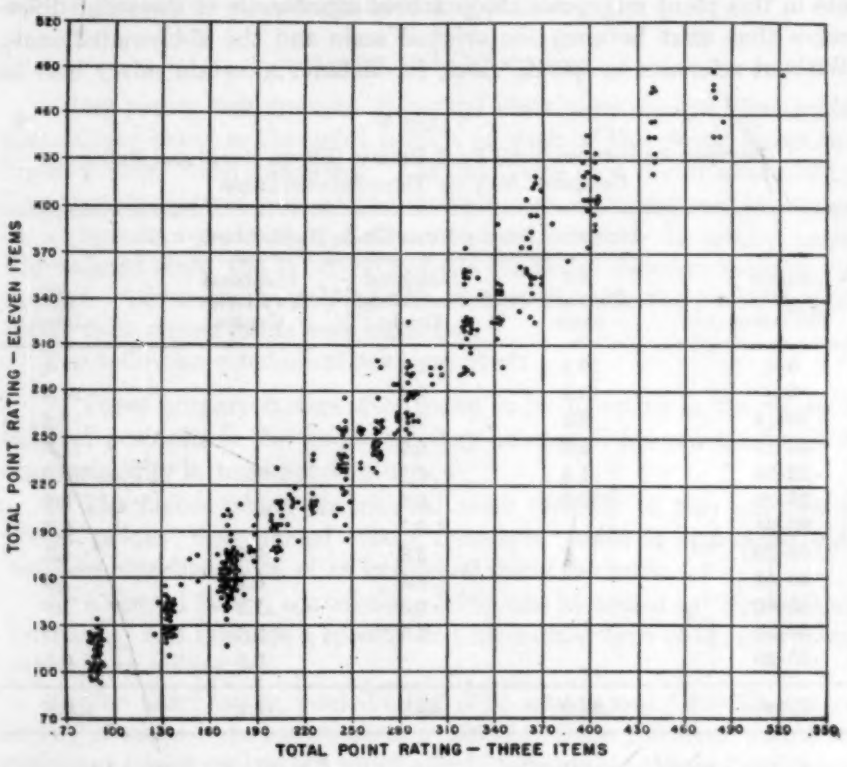


FIG. 2. Scattergram showing ratings computed from the three item scale plotted against total point ratings (all 11 items) for 391 salaried positions.

used in this plant, the rated jobs are grouped into labor grades, each having an interval of 30 points. These labor grades are indicated in the scattergram by the grid lines. Thus all jobs which were rated under 100 points fall in labor grade 1, those which received between 100 and 130 points in grade 2, and so.

Labor Grade Displacement. The deviations of the computed values for the different jobs and the number of labor grades displaced are presented in Table 5. Of the jobs studied, 44% remained in the same labor grade, 53% were displaced by one labor grade, and only 3% were dis-

placed by two labor grades. The table also shows that only 16% of all the jobs, deviate from their original rating by more than 30 points, the range of one labor grade, or the point difference between the lowest and highest rated job in any labor grade. The results obtained indicate a high degree of comparability between the two scales.

Wage Structure Considerations. An examination of the salary schedule in this plant minimizes the practical significance of the point differences that exist between the original scale and the abbreviated scale. Without reference to specific rates, for instance a certain salary may be

Table 5

Discrepancies Between Total Point Ratings (Eleven Items) and Ratings Computed from the Three Selected Items

Points of Deviation	Percent of Jobs by Labor Grade Displacement			
	No Displacement	Displaced One Labor Grade	Displaced Two Labor Grades	All Jobs
0–4	16.1	1.3		17.4
5–9	13.1	8.4		21.5
10–14	8.4	9.2		17.6
15–19	4.6	6.9		11.5
20–24	1.5	7.5		19.0
25–29	0.3	6.6		6.9
30–34		6.7		6.7
35–39		3.8	0.5	4.3
40–44		1.5	0.3	1.8
45–49		0.5	1.3	1.8
50–54		0.5	0.5	1.0
55–59			0.5	0.5
Totals	44.0	52.9	3.1	100.0

paid at one time or another to employees on jobs ranging through several different labor grades. This fact tends to reduce the importance of a few points difference between the two scales.

The Reliability of Point Rating. The reliability of the original ratings is not known. It would be impractical to obtain re-ratings by a group of raters of equal ability and familiarity with the jobs. Therefore it is impossible to determine whether or not the correlation between the original and the abbreviated scales is as high as the reliability of the scale itself. Lawshe [7] provides some enlightening data on the reliability of point ratings in an industrial plant. The same jury of supervisors and analysts rated a group of five jobs at three different intervals within a period of

[7] C. H. Lawshe, Jr. *Op. cit.*

about three weeks. Their ratings fluctuated anywhere from 25 to 100 points, the average change being 71 points. The relative unreliability of human judgment further diminishes the practical importance of the differences between the two scales. The results yielded by this study reveal a high degree of similarity and agreement between the original and the abbreviated scales; therefore for all practical purposes, the two scales can be considered almost identical in operation.

Summary and Conclusions

Salary rating data from an industrial plant were obtained and intercorrelations between the point ratings on each of the eleven items and "total points" were computed. The data were then factor analyzed by Thurstone's centroid method and the values were transformed or rotated by a procedure outlined by Peters and Van Voorhis. In setting up an abbreviated scale, the Wherry-Doolittle shrinkage selection method was applied. Three items yielding the highest R with the "total point" rating from eleven items were identified.

The following conclusions are supported:

1. Three primary factors were found to be operating in the rating of salaried positions in the plant and they accounted for practically all of the variability in total point ratings.

2. The factor which contributed most (95.6%) of the variance in "total points" was named "Skill Demands" since it represents skill requirements demanded of an individual doing the work.

3. A second factor, contributing 3.7% was identified as "Supervisory Demands" and includes a specific and specialized type of skill involving supervisory ability.

4. The third factor, contributing .7%, was named "Job Characteristics" because it embraces certain aspects of the job over which the employee has no control but which affect his mental or physical well being.

5. The factors identified are similar to those reported by Lawshe and Satter for hourly-paid jobs in three industrial plants. "Supervisory Demands," however, did not appear in that study since this factor is less characteristic of hourly paid jobs.

6. The items selected for an abbreviated scale were "experience," "complexity of duties," and "character of supervision." The multiple R between the three items and "total points" was .980 and the standard error of estimate 20.3.

7. About 96% of the variance in "total points" may be attributed to the three selected items; the other eight items in the scale contributed the remaining 4%.

8. If the abbreviated scale were used in this plant, 44% of the salaried positions would remain in the same labor grade, 53% would be displaced one labor grade, and 3% would be displaced two labor grades.

9. The practical significance of the differences which exist between the point ratings on the original scale and those on the abbreviated scale are minimized by the flexibility of the plant salary scale and the probable unreliability of the ratings.

10. An abbreviated scale consisting of about three items would probably give results practically identical to those obtained by the longer and more complicated eleven item scale, and would greatly reduce the time required to do the rating.

Received April 19, 1945.

A Preliminary Report on the Miami-Oxford Curve-Block Series *

P. L. Mellenbruch

Department of Psychology, University of Kentucky

The Miami-Oxford Curve-Block Series is the outgrowth of a number of years of practical experience during which the writer served as the director of a psychological clinic and was engaged in personnel work in industry.[1] Out of this experience there appeared to be need for a manipulative test involving form perception which would reveal a person's ability and his persistence in solving complex spatial problems. In the early days of his clinical work the writer designed a single curve-block which proved helpful. It is somewhat similar to the O'Connor Wiggly Block [2] which appeared later. It was felt, however, that a series of such blocks arranged in steps of increasing difficulty would be of much more value.

In the production of the curve-block series the following principles were kept in mind: (1) The blocks in the series should range from very simple to quite difficult; (2) the steps of increasing difficulty from one block to the next should be approximately equal; (3) the total series should not be too cumbersome but should be large enough for the smallest parts to be readily perceived and easily handled; (4) the series should be so devised as to be practically self-administering; and (5) the test should reduce to a minimum the factor of chance success or failure.

The Miami-Oxford Curve-Block Series consists of six rectangular blocks of the same over-all dimensions. Each block is cut lengthwise along curved lines so as to form a number of irregular sections. All of

* Acknowledgment is hereby given to Dr. E. F. Patten, Chairman of the Department of Psychology, Miami University, for his encouragement and helpfulness in developing the Curve-Block Series and the data here presented. The following students of Miami University gave generously of their time in the construction of the blocks and the collection of the data for this report: Robert Gehlker, Dorothy Hoffmeister, Bettie Perkins, and Robert Dixon.

[1] In 1926 the writer devised and began using the curve-block, shown in the foreground of Figure 1, in his testing of court cases and social service agency cases. This original block, somewhat smaller in size than the blocks of the present series, is cut lengthwise so as to form fifteen irregular sections. It is intended for adults and older youths and was used to help estimate the manner in which they would attack a rather difficult concrete problem as well as to indicate their persistence.

[2] O'Connor Wiggly Block, Johnson O'Connor (1928); Stevens Institute.

the parts of these blocks are so designed that no inside surface is flat or straight. In no case are the sections in any one block identical or reversible, nor are the sections of different blocks interchangeable.

The succeeding blocks in the series were made successively more difficult by increasing the number of cuts, hence the number of parts, in each succeeding block and also by modifying the shape of the sections in the various blocks of the series. The extent to which this has achieved the desired results can be seen by reference to the data below.

The six curve-blocks are each of a different color and are referred to in this report according to color. The *blue* block consists of three parts,

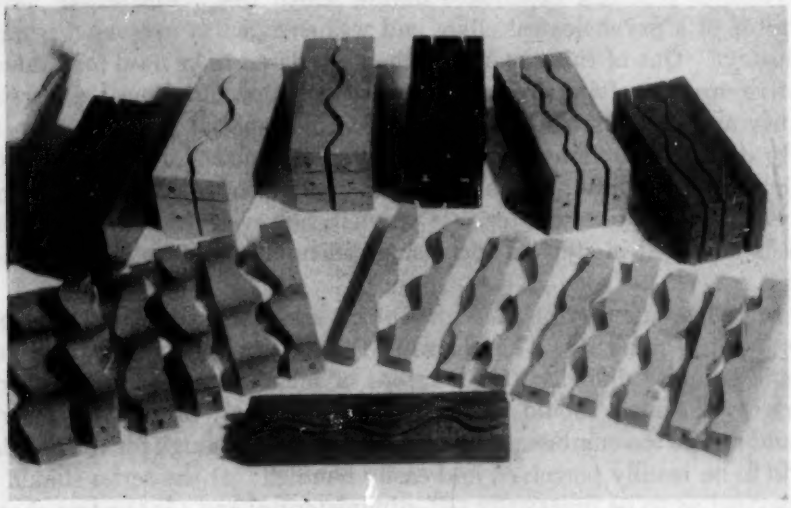


FIG. 1. Miami-Oxford Curve-Block Series (with original block in foreground). Two blocks shown disassembled.

the *white* of four parts, the *green* of six parts, the *black* of eight parts, the *orange* of nine parts, and the *red* of twelve parts (see Fig. 1).

As a result of experimental administration of the series and the analysis of preliminary data, a number of changes were made in the size of the blocks and in the shape of their parts. The original blocks were three and a half inches square by nine inches long. These were found to be unnecessarily large so the present size was adopted, the blocks now being approximately three inches by three inches, and eight inches long. Some changes also in the shape of certain parts were made to alter their relative difficulty and to prevent sections, once assembled, from falling apart or from presenting cues for ready solution. The present series appears to be quite free from these objections.

Administration

The directions and procedures for administration are simple and are followed uniformly. Each block in the series is presented one at a time to the examinee. As the first assembled block is presented, the examiner says: "I am going to take this block apart and when I say 'Go,' I want you to put it together as quickly as you can." Then at an easy comfortable rate, the examiner disassembles the block, putting the pieces in a fixed order in front of the examinee. This is readily accomplished because the parts of each block are numbered according to this sequence.

When the examinee has reassembled the first block, each subsequent block is presented already assembled and then taken apart with the sections placed in the prescribed order. The instructions for the blocks after the first one are: "Put these pieces together as fast as you can." Such simple directions prove to be quite adequate in view of the fact that the examinee is given the easiest block first.

In this preliminary study, both the time required and the number of errors made in putting the sections of each block together were recorded. The time in each instance was accurately obtained by means of a stop watch. An error was counted if any two blocks which did not belong together, or which did not belong together in the manner attempted, were actually brought into contact with one another. Each new contact in putting the blocks together incorrectly was counted as an error even though the same error had been made before. Subsequent findings may show that the "time" score is adequate for most purposes, in which case no record need be kept of the "errors" made.

Results

It should be kept in mind that the data presented here are given in the nature of a preliminary report. The results, however, show trends which should for the most part, be borne out by subsequent findings. Further validation is planned, particularly with reference to the use of this Curve-Block Series in industry.

The data in Tables 1, 2, 3, and 4 were obtained from 93 cases, using the Miami-Oxford Curve-Block Series in its final revised form. From Table 1, it will be seen that the mean time increases from one block to the next. The proportional increases shown in Table 2 were obtained by dividing the mean time for each block by the mean time of the next easier block. This indicates, for example, that on the basis of present data the white block is 2.63 times as difficult as is the blue block; that the green block is 1.86 times as difficult as is the white block, and so forth. Stated somewhat differently we read from the second column in Table 2 that the blue block is 38% as difficult as the white block, that the white block is

55% as difficult as the green block, and so forth. These percentages of increase are obtained by dividing the mean time of each block by the mean time for the next more difficult block in the series.

While the steps of increasing difficulty are not as equal as desired, yet they are reasonably regular and may possibly prove to be more so from

Table 1

Means and Sigmas of Time Scores (Time in seconds). $N = 93$

Block	Score			Mean	σ_M	σ	σ_σ
	High	Low	Range				
Blue	2	74	72	16.0	1.30	12.56	.92
White	5	237	232	42.2	3.99	38.49	2.82
Green	12	470	458	77.5	6.99	67.43	4.94
Black	25	1018	993	170.6	16.26	156.81	11.46
Orange	71	885	814	269.3	19.23	185.45	13.60
Red	52	1730	1678	395.3	26.67	257.10	18.85
Total*	312	2735	2423	969.7	52.29	504.25	36.96

* "Total" represents the individual's score on the entire series of blocks.

further data. That these differences between the succeeding blocks of the series are significant (in terms of the "time" scores) is quite evident from the critical ratios [3] given in Table 2. For the most part they are well above the requirements of real differences.

Table 2

Relative Difficulty of Blocks on Basis of Time (Time in seconds). $N = 93$

	Proportional Increase	Percentage of Increase	Critical Ratio
Blue-white	2.63	.38	6.25
White-green	1.86	.55	4.35
Green-black	2.20	.45	5.26
Black-orange	1.52	.64	3.92
Orange-red	1.47	.68	6.55

Table 3 presents the data from these same 93 cases in terms of *errors*. It will be seen that in errors as well as in time the blocks are graded in difficulty. The extent of such differences in difficulty is presented in Table 4, where it is indicated that the blocks increase rather regularly in difficulty up to the orange and red blocks, at which point the difference is not so great. And, referring to Table 2, it is seen that on the basis of

[3] The critical ratio here referred to represents the difference between the means of any two blocks divided by the square root of the sum of the squares of the standard error of each mean.

mean time there is also less difference between the orange block and the red block than between other blocks of the series. The critical ratios in Table 4 show that the succeeding pairs of blocks in the series differ signifi-

Table 3

Means and Sigmas of Error Scores. $N = 93$

Block	Score			Mean	σ_M	σ	σ_σ
	Low	High	Range				
Blue	0	14	14	3.7	.35	3.34	.24
White	0	22	22	6.1	.46	4.44	.33
Green	0	60	60	10.7	.98	9.42	.69
Black	0	75	75	17.6	1.79	17.23	1.26
Orange	2	142	140	32.0	2.67	25.73	1.88
Red	1	215	214	40.3	3.82	36.80	2.69
Total*	8	451	443	109.7	7.98	76.98	5.72

* "Total" represents any individual's score on the entire series of blocks.

cantly in terms of errors except for the orange and red blocks. Here there are only 96 chances in a hundred that the differences between these two means will hold. From Table 2, however, it is evident that these two blocks present significant differences on the basis of time.

Table 4

Relative Difficulty of Blocks on Basis of Errors. $N = 93$

	Proportional Increase	Percentage of Increase	Critical Ratio
Blue-white	1.68	.60	4.28
White-green	1.74	.57	4.21
Green-black	1.64	.60	3.38
Black-orange	1.82	.55	4.50
Orange-red	1.26	.79	1.77

Correlations within the curve-block series itself show the following relationships between *time* and *errors* for the different blocks and for the series as a whole:

It will be seen from Table 5 that the correlation between time and errors is rather high on four of the six blocks. This probably means that

Table 5

Coefficient of Correlation between Time and Errors. $N = 93$

	Blue	White	Green	Black	Orange	Red	Total
Correlation	.74	.72	.63	.77	.27	.54	.64
P.E.	.032	.030	.042	.028	.065	.052	.041

a time score on these four easiest blocks of the series is sufficient. However, in the light of rather low correlations between time and error on the orange and red blocks it might be of advantage to record both time and errors separately for the last two blocks of the series. This would reveal those who, in a given time, make many hurried attempts at solution as opposed to those who are more deliberate in their procedures.

The intercorrelations given in Table 6, are for the most part low enough to indicate that the blocks are sufficiently dissimilar to justify their inclusion in the series. These may indicate unreliability, however, and therefore indicate the need for a series of blocks. On the whole,

Table 6

Intercorrelations*

	White	Green	Black	Orange	Red
Blue	.39	.11	.28	.27	.32
	White	.12	.11	.09	.22
		Green	.55	.34	.19
			Black	.42	.31
				Orange	.43

* Odd (1, 3, 5) vs. Even (2, 4, 6) $r = .59 \pm .045$; corrected Spearman-Brown $r = .75$.

there appears to be more agreement between adjacent blocks in the series than between those removed farther from one another. While the odd-even correlation is not as high as might be desired, it still is significant. A further check on the reliability of the series by a considerable number of re-tests is planned.

It is possible that the Miami-Oxford Curve-Block Series could be used to advantage as a self-administering test. This might be done by placing the six blocks, each one disassembled with the segments in the proper order, on a long table and telling the subject to reassemble them as quickly as possible. The total time required for the series could be used as the score. Or a time limit of ten minutes or so might be used, in which case the score would represent the number of whole blocks reassembled, plus parts correctly placed in the last block attempted but not completed.

Summary

Altogether it appears that the Miami-Oxford Curve-Block Series will meet a particular need in the field of measurements, especially in industrial and clinical situations. It further appears that this series has largely met the requirements of simple administration, convenient size, steps of approximately equal difficulty, and the control of chance factors of success and failure.

Received April 13, 1945.

A Comparative Study of Three Tests for Color Vision *

Harriet Foster

University of Minnesota

Because of the importance of color blindness tests in all types of personnel work, the question of the effectiveness and the comparability of such tests is important. To the author's knowledge, only two studies have been published comparing various tests for color blindness. R. B. Philip compared the Edridge-Green, Nagel, Holmgren wools, Philip color perception, and Ishihara tests; correlations ranged from .48 to .79 for color-blind and anomalous trichomats and from .50 to .90 for all cases of defective color vision.[1] G. J. Thomas compared the American Optical Company's Pseudo-Isochromatic Plates, Ishihara, Farnsworth-Munsell, and Color aptitude test designed by the committee of the Inter-Society Color Council. Correlations were computed for color-blind and normal subjects separately, but not for the group as a whole. The most significant of these coefficients was between normal subjects on the Ishihara and the American Optical Company tests. Correlations between the various tests ranged from .14 to .64.[2] They show, however, only agreement *within* the classifications of color-blind and normal.

The important problem in personnel work is whether or not the tests are in agreement in their diagnosis of color-blind and normal. This can best be revealed by scatter tables which center attention upon specific instances of agreement and disagreement.

The three tests used in the present study were: (1) Pseudo-Isochromatic Plates put out by the American Optical Company, (2) Tests for Color Blindness by Ishihara, and (3) Tests for Color Blindness by Milton B. Jensen. The Pseudo-Isochromatic Test consists of 48 plates. Each plate has either one, two, or three numbers, a letter, or a curved pathway on it in a different color or a different shade of color from the background. The figures and background are in colored dots rather than a solid color. Thirteen plates of the Ishihara Test were used. They are all numbers and quite similar to the Pseudo-Isochromatic Plates. This is a well

* This study was conducted in the Psychological Laboratory, University of Minnesota, under the direction of M. A. Tinker and D. G. Paterson.

[1] Philip, R. B. A comparison of color-blind tests. *Amer. J. Psychol.*, 1938, 51, 482–488.

[2] Thomas, G. J. Visual sensitivity to color: A comparative study of four tests. *Amer. J. Psychol.*, 1943, 56, 583–591.

known test and has the disadvantage that some people have been able to obtain and memorize the normal responses. The test by Jensen has four plates, all of which are made to look like faces of clocks. The subject reports how many hands he sees and what hours they point to. The brevity of this test would be its chief advantage. All three of the tests have at least one plate which the color-blind and normal people see alike. This is to detect malingerers. The other plates are supposed to get different responses from color-blind subjects than from persons with normal color vision.

Each of the three tests was given according to the standard directions accompanying each test. The Pseudo-Isochromatic Plates were held approximately three feet from the subject and about level with his eyes. The Jensen test was held about ten feet from the subject unless his vision was poor. In that case the chart was held close enough so that the subject could see lines in the center of it. The Ishihara was either held at about three feet from the subject or laid out on a table with the subject standing.

Two hundred men, ranging in age from 17 to 56, acted as subjects. Women were not used because such a low percentage of them are color-blind.

The following method of scoring was used.[3] Two points were counted for each color-blind response, nothing was counted for normal responses. If a number was misread, such as calling a "3" an "8" or a "7" a "2," one point was taken off since this seems to show some deficiency in color vision. If an apparently normal subject saw the number on a plate which only color-blind persons are supposed to see, one point was counted since this too seems to show a slight deficiency. The highest score possible on the Pseudo-Isochromatic Plates was 86, on the Ishihara it was 24, and on Jensen's test 6.

Table 1 shows the arbitary classifications made of the subjects as normal, borderline, or color-blind on the three tests. More than the usual percentage of subjects were designated as color-blind by the Pseudo-Isochromatic Plates and the Ishihara test. The reason for this is probably that the method of obtaining subjects tended to attract more color-blind subjects than would be found in a random sample. Although the percentages of normal and color-blind subjects are in good agreement for the American Optical Company test and the Ishihara plates, there is still the question of whether the same subjects are color-blind on the two tests. It is also necessary to compare the Jensen test with the other two tests.

[3] This method was used by Thomas, *op. cit.*

Table 1

Classifications of Responses Made on Three Tests of Color Blindness. $N = 200$ men

Classifications	Amer. Opt. Co. Pseudo-Isochromatic Plates				Ishihara Plates				Jensen Test			
	Range of Scores	Mean Score	*N*	%	Range of Scores	Mean Score	*N*	%	Range of Scores	Mean Score	*N*	%
Normal color vision	0–17	2.5	177	88.5	0–5	.4	178	89.0	0–2	.5	189	94.5
Borderline	18–41	26.3	3	1.5	6–11	8.2	4	2.0		—	—	—
Color-blind	42–71	58.2	20	10.0	12–24	20.4	18	9.0	3–6	4.5	11	5.5
Total	0–71	8.5	200	100.0	0–24	4.3	200	100.0	0–6	.7	200	100.0

Scatter tables were made for each pair of tests. Table 2 shows the degree of agreement between the Ishihara and the Pseudo-Isochromatic Plates. There is some disagreement on the borderline cases, but none as serious as to diagnose a person as normal on one test and color-blind on the other. Two cases are borderline on the Ishihara and color-blind on the Pseudo-Isochromatic, two are borderline on both tests, and one is normal on the Ishihara and borderline on the Pseudo-Isochromatic. The other 195 cases agree in the classification of color-blind and normal. This is a 97.5% agreement.

Table 3 gives the results for the Ishihara and Jensen tests. It is quite apparent that they do not agree on deviate cases. Because of the limited range of scoring units on the Jensen test, it did not seem worth while to designate any cases as borderline. It was difficult even to distinguish between normal and color-blind subjects. None of the subjects diagnosed as color-blind by the Jensen test was found to be normal on the Ishihara and only two of them were borderline cases on the Ishihara. But nine cases diagnosed as normal by Jensen were color-blind on the Ishihara and two were borderline. There were 187 cases in agreement in the classification of color-blind and normal. A classification of color-blind on the Jensen test is, therefore, much more valid than one of normal if the Ishihara classification is assumed to be correct. In other words, the Jensen test results may be accepted if a person tests as color-blind, but it "misses" so many cases that its value is greatly reduced.

Table 4 shows the relationship between the Jensen and the Pseudo-Isochromatic Plates. Of those testing color-blind on the Jensen, none was found to be normal by the Pseudo-Isochromatic Plates and only one was a borderline case. But again those testing as normal on the Jensen included cases which tested as borderline or color-blind on the Pseudo-Isochromatic Plates. Ten of the Jensen normal subjects are color-blind according to the Pseudo-Isochromatic Plates and five are borderline. The remaining 184 cases agree. As in Table 3, we find the Jensen test good on the subjects testing as color-blind, but not for those diagnosed as normal.

Table 5 shows the actual scores made by the 23 deviate cases. In this way it is possible to compare the three tests simultaneously. It can be seen again that the Jensen test is not in close agreement with the other two tests on these cases. Nine cases diagnosed as color-blind by the Ishihara and Pseudo-Isochromatic Plates were classified as normal on the Jensen test. Three of them actually had perfect scores on the Jensen test. Although none was found normal on the Pseudo-Isochromatic and Ishihara and color-blind on the Jensen, there is one case diagnosed as borderline on these two tests and color-blind on Jensen. The Ishihara

Table 2

Scatter Diagram Showing Scores Made on Ishihara and Pseudo-Isochromatic Plates. $N = 200$ men

Ishihara Test Scores and Diagnoses		Pseudo-Isochromatic Plate Scores with Diagnoses												
		Normal			Borderline				Color-blind					
		0–5	6–11	12–17	18–23	24–29	30–35	36–41	42–47	48–53	54–59	60–65	66–71	*N*
Color-blind	24										1	1	3	5
	22–23										2	1	1	4
	20–21									1	1	1		3
	18–19								1		1			2
	16–17									1	1			2
	14–15										1			1
	12–13										1			1
Borderline	10–11						1			1				2
	8–9													0
	6–7				1						1			2
Normal	4–5	2		1										3
	2–3	16	3	1										20
	0–1	139	14	1		1								155
	N	157	17	3	1	1	1	0	1	3	9	3	4	200

Table 3

Correlation Between Jensen and Ishihara Tests. $N = 200$ men

Jensen		Ishihara: Normal 0–1	2–3	4–5	Borderline 6–7	8–9	10–11	Color-blind 12–13	14–15	16–17	18–19	20–21	22–23	24	*N*
Color-blind	6													2	2
	5											1		3	4
	4				1					1			1		3
	3						1						1		2
Normal	2	34	5	1			1	1			1		1		44
	1	2	1		1				1			1	1		7
	0	119	14	2						1	1	1			138
	N	155	20	3	2	0	2	1	1	2	2	3	4	5	200

Table 4

Correlation Between Jensen and Pseudo-Isochromatic Plates. $N = 200$ men

Jensen		Pseudo-Isochromatic Plates: Normal 0–5	6–11	12–17	Borderline 18–23	24–29	30–35	36–41	Color-blind 42–47	48–53	54–59	60–65	66–71	*N*
Color-blind	6											1	1	2
	5									1	1		2	4
	4									1	1		1	3
	3						1					1		2
	2	36	3	1						1	3			44
Normal	1	3			1						3			7
	0	118	14	2		1			1		1	1		138
	N	157	17	3	1	1	1	0	1	3	9	3	4	200

test seems to fall down on case 10. R. E. V. This subject was classified as color-blind by the other two tests and borderline on the Ishihara. The author can offer no explanation for this. In general, it is clear that the Ishihara and the Pseudo-Isochromatic tests are in close agreement.

In testing, the experimenter found one thing that could be improved on in the Jensen test. The third plate was failed by 40 of the normal subjects. Many of the subjects said that they were confused by it because there was one curved line and one straight line. Many failed to

Table 5

Results of Three Color Blindness Tests for the 23 Deviate Cases
(Color-blind or borderline on any one test)

	Pseudo-Isochromatic Test		Ishihara Test		Jensen Test	
Individual	Score	Diagnosis	Score	Diagnosis	Score	Diagnosis
1. K. M.	71	C. B.	22	C. B.	4	C. B.
2. R. T. M.	70	C. B.	24	C. B.	5	C. B.
3. L. D. H.	68	C. B.	24	C. B.	5	C. B.
4. D. N. L.	66	C. B.	24	C. B.	6	C. B.
5. S. H.	65	C. B.	24	C. B.	6	C. B.
6. W. L. K.	63	C. B.	22	C. B.	3	C. B.
7. L. H. M.	61	C. B.	20	C. B.	0	Norm.
8. P. F.	59	C. B.	24	C. B.	5	C. B.
9. M. P. S.	59	C. B.	22	C. B.	2	Norm.
10. R. E. V.	58	C. B.	6	Bord.	4	C. B.
11. R. W. C.	56	C. B.	21	C. B.	1	Norm.
12. C. B. H.	56	C. B.	12	C. B.	2	Norm.
13. E. L.	55	C. B.	22	C. B.	1	Norm.
14. D. M.	55	C. B.	22	C. B.	1	Norm.
15. J. A. H.	54	C. B.	17	C. B.	0	Norm.
16. S. S. S.	54	C. B.	15	C. B.	1	Norm.
17. A. J. B.	53	C. B.	21	C. B.	5	C. B.
18. J. B.	49	C. B.	16	C. B.	4	C. B.
19. M. E. E.	48	C. B.	10	Bord.	2	Norm.
20. C. N.	44	C. B.	19	C. B.	0	Norm.
21. L. O.	35	Bord.	11	Bord.	3	C. B.
22. C. P.	26	Bord.	0	Norm.	0	Norm.
23. D. P.	18	Bord.	6	Bord.	1	Norm.

report the straight line because they were confused by the curved one. Even if this difficulty were removed, there would still be disagreement on the subjects diagnosed as normal by Jensen. Apparently a 3 or 4 plate test is too unreliable for individual diagnosis.[4]

[4] The results of this study suggest, but do not prove, that a color blindness test consisting of a small number of plates may not be valid.

If the Ishihara test is accepted as a valid test of color blindness, as is customary, the Jensen test is of doubtful value in the diagnosis of color blindness. Possibly the latter is too short or perhaps the kind of response required is not as good as that of the other two tests. It might be just as ineffective with 15 plates. The only thing against the Ishihara test is its widespread use and the possibility that coaching or memory from previous exposure may invalidate the results. The Pseudo-Isochromatic Plates take more time, but it might be possible to shorten it if the only thing desired is to distinguish between the normal and color-blind. Or it may be necessary to have a large number of plates in order to be certain to identify all color-blind cases. Further research is indicated.

Summary

1. Few studies have been published comparing tests of color blindness. Because of their widespread use, it is desirable to know how the scores on the different tests agree.

2. Two hundred men were used as subjects. Each man took the Ishihara, the Pseudo-Isochromatic, and the Jensen tests.

3. The results showed that the Ishihara and Pseudo-Isochromatic Plates were in close agreement with each other, while the Jensen showed far less agreement with either of the other tests. The Jensen test mistook nine color-blind subjects for normal. If the Ishihara test is accepted as valid, the Jensen test does not have adequate validity for individual diagnosis.

Received March 5, 1945.

Studies in the Application of Motor Skills Techniques to the Vocational Adjustment of the Blind

Mary K. Bauman

Personnel Research and Guidance, Philadelphia, Pa.

The work here presented reports the application of psychological measures to the prediction of success in industrial employment for blind persons. The Trainee Acceptance Center was established to give vocational guidance to any adult who wished it in connection with training for or placement in war work. However, when the first *blind* person applied for advice, neither the experience of staff members nor the literature of vocational guidance provided adequate techniques for determining the employability of this handicapped group. Applications of psychology to the educational and more general problems of the blind, as presented in Hayes' *Contributions to a psychology of blindness*,[1] gave basic material. Early efforts in vocational guidance are reported by the same author in a later publication.[2]

Our attack upon the guidance of persons with deficient vision involved the measurement of skills required and consideration of the results of this measurement in the light of the training, environmental, and personality factors in each particular case. It seemed reasonable to suppose that this procedure would also yield satisfactory results if applied to blind clients.

Plan of Research

The first step was to find tests suited to measuring the skills a blind person could use on an industrial job. The fact that the Center had been established to guide individuals seeking war work oriented our early studies with the blind purely toward industry, not toward vending stands, social work, or other occupations open to them. The second step was to establish quickly the rough standards a blind person should meet on these tests if he was to be successful in industry. This was accomplished with the cooperation of a group of blind persons already employed on production who subjected themselves to the testing. It seemed possible how-

[1] Hayes, Samuel P. *Contributions to a psychology of blindness.* New York: American Foundation for the Blind, 1941.

[2] Hayes, Samuel P. History and present status of aptitude tests for the blind. *Unpublished Manuscript.*

ever that the learning of blind persons, particularly in handling concrete material, was so different from the learning of seeing persons that the results of two hours of testing could not be relied upon. Therefore, with the assistance of the Pennsylvania Institute for the Instruction of the Blind, the learning curves of blind persons dealing with concrete material were studied. Finally, even if the tests did appear to establish a useful standard for industrial placement, it was possible that we were merely measuring in several ways a single ability or group of abilities; a fourth step therefore involved the intercorrelation of the results of the various tests.

Tests Adapted for Use With Blind Clients

1. Minnesota Rate of Manipulation.[3]
   - *A.* Displacing. The first part of the Rate of Manipulation test was modified. The blind subject moves the blocks around the board in a pattern rather than following the standard procedure known as "Placing." Three practice and three test trials are given.
   - *B.* Turning. The second part of the Rate of Manipulation test was used without change except that again three practice and three test trials were given.
2. Pennsylvania Bi-Manual Worksample.[3]
   - *A.* Assembly. For the person who was totally blind or had only light perception, a preliminary trial was given in which he was asked to assemble a nut and bolt and throw them into a box just beyond the board. For persons able to distinguish the holes in the board this preliminary trial was omitted. In all cases, two trials involving placement of the assembled nuts and bolts into the holes in the board were given, and the final score used was the better of the two. On these two trials no modification was made except that the blind person was encouraged to use his dominant hand to locate the hole into which he should drop the nut and bolt.
   - *B.* Disassembly. This portion of the Worksample was used without modification but, again, was administered twice and the better of the two trials was used as a final score.
3. Toolsample.

   This is a worksample in which various kinds of nuts and bolts are removed from one board and tightened in another with the aid of tools. Results reported in this paper were obtained by its use with male clients only; it was rarely used with female clients. Its complexity as compared with motor skills tests, and its similarity to many real jobs, make it particularly important in the evaluation of clients who have had trade experience before the loss of vision. Scoring allows for training on the first bolt if the testee cannot remove it by himself. During the remainder of the test it is possible to observe qualitatively and measure quantitatively his ability to profit by this training on the first bolt and by the experience which each succeeding bolt offers. It can therefore be used as a measure of the trainability of the individual with this type of material within the limits of a single session.

[3] Descriptions of Minnesota Rate of Manipulation and Pennsylvania Bi-Manual Worksample, including norms and instructions for their use with the blind, may be obtained from the publisher, Educational Test Bureau, Minneapolis, Minn.

Tests Tried but Discarded

1. O'Connor Finger Dexterity.
   This test provides useful information in the cases of persons having better than light perception but required too accurate orientation in space for our totally blind clients.
2. Purdue Pegboard.
   The above statement also applies to this test.
3. Object Sorting and Assembly.
   In the attempt to reproduce the fundamentals of certain jobs, a test was devised which required the sorting into separate boxes of five small objects easily distinguished by touch, and their subsequent assembly into a pattern on a stick. However, even those blind persons who performed well on other tests and who had been successful in industry could not make scores within the normal range of seeing persons on this exercise, despite additional practice. We were forced to conclude that vision gave advantages in the performance of this task which could not be overcome by any reasonable amount of practice for even the more able blind clients.

Additional Data for Vocational Guidance

A measure of general mental ability was included in every examination. For this the Wechsler-Bellevue Verbal Scale, the Interim Hayes-Binet, and the Kent E-G-Y Test were used, the test to be used in a specific case being dictated by the background and special problem of each client.

A very important part of the results in every case is the qualitative observation made by staff members during the testing. Ease in following directions, freedom from confusion, amount and kind of supervision needed are all noted. Coordination, orientation in the work space, distractibility, and motivation are also described. Staff members are encouraged to write complete descriptions of everything they observe.

A medical examination paralleling those given by industrial firms is included in the program of all clients at the Trainee Acceptance Center.

Subjects

As defined by law, blindness includes all persons having up to 20/200 vision. Our work quickly demonstrated that in the presentation of test results subdivisions of the visually handicapped would be desirable. Results of clients who are totally blind or have so little light perception as to be of no value in the test situation are presented together; this group will be referred to as "Totally Blind." Results of clients having more than light perception but not more than 10/200 vision will be presented together, and will be referred to as "10/200." Results of clients having better than 10/200 vision but not more than 20/200 will be presented together, and referred to as "20/200." In each case the notation indicates the best vision of the client, whether with one or both eyes, and with such aid as glasses could give. A division by age into a group under 40 and a group 40 and over was made because of the experience that industry favors the employment of the younger group and because all norms used at the Center are based on test results of persons under 40 years of age. Table 1 shows the distribution of subjects in each group.

Thus a total of 312 legally blind persons was studied, 231 of whom were less than 40 years of age. In this we are extremely fortunate because the preponderance of older persons in the total blind population

Table 1

Analysis of Subjects in Terms of Age and Degree of Blindness

Group	Under 40 Years of Age			40 Years of Age and Over		
	Men	Women	Total	Men	Women	Total
A. Totally blind	85	43	128	39	9	48
B. "10/200"	31	19	50	17	5	22
C. "20/200"	25	9	34	7	4	11
D. Employed group (totally blind)	18	1	19			
Total	159	72	231	63	18	81

makes it difficult even in many cities to reach a significant number of blind persons of employable age for study.

Test Results

For greater ease in handling results where sex differences are known to exist, and where raw scores can have little meaning for persons unfamiliar with the tests, all results in this section are presented in standard scores based on the testing of hundreds of seeing clients.

Tables 2, 3, and 4 give the number of persons under 40 years of age attaining each standard score, subdivided by test and amount of vision. A line is inserted between the standard score of 4.0 and scores below. This was done because this is generally considered at the Center to be the cut-off point in any trait; those with scores below this line are handicapped in any competitive situation where this trait is required. A direct positive relationship is evident between success in all six tests and amount of vision. Group C (20/200 vision) shows the largest per cent with a standard score above 4, Group B (10/200 vision) has the next largest per cent, and Group A (totally blind) has the smallest per cent. Total blindness is of the greatest disadvantage on the Displacing Test, where only 18% of the group attain scores within the normal range. It is of least disadvantage with the Turning Test where orientation in space is at a minimum.

There is in general a greater piling up of scores in the lower 2%, or standard score of 3.0 or less, than would be expected of the normal seeing population; this is particularly true with the Displacing Test. For the 10/200 and 20/200 groups, a number of individuals close to normal expectance obtain scores better than 6.0. For the totally blind group,

Table 2

Distribution of Standard Scores of Totally Blind Men and Women (Group A)

Standard Score	Displacing	Turning	Assembly	Disassembly	Toolsample, Part 1	Toolsample, Part 2
7.0		1				
6.75		1				
6.5						
6.25					2	
6.0		2		1		
5.75		2		1	1	
5.5		5	1	2		
5.25		2	1	5	3	
5.0		10	3	6	1	
4.75	1	10	3	7	3	
4.5		13	3	6	1	3
4.25	3	13	6	12	7	1
4.0	2	11	16	7	6	7
3.75	2	10	13	11	5	5
3.5	6	3	10	13	9	9
3.25	5	13	11	24	12	4
3.0	4	7	18	9		8
Less than 3.0	50	25	43	24	1	15
N	73	128	128	128	51	52

scores are concentrated below the standard score of 6.0 except for the Turning Test and Toolsample No. 1, demonstrating the importance of orientation in space on all the other tests.

Results for persons over 40 years of age, not shown in tabular form, indicate, roughly, only a little better than half as much chance that the older clients will obtain acceptable scores. There is also greater concentration of results below the standard score of 6.0; not only is there less chance of an acceptable score, but very much less chance of an exceptional score. However, no age is an absolute cutoff. One of the best records was made by a man of 54.

Since differences in amount of vision cause such significant differences in test performance and our major interest is in obtaining tests for clients not aided by vision, the remainder of this study concerns only the Totally Blind group. In addition, all further figures are based on the results from persons under 40 years of age.

Table 3

Distribution of Standard Scores of Men and Women with Better than Light Perception but not More than 10/200 Vision (Group B)

Standard Score	Displacing	Turning	Assembly	Disassembly	Toolsample, Part 1	Toolsample, Part 2
7.0		1				
6.75			1	1	1	
6.5		1	1			
6.25		1		2		
6.0	1	1		1		
5.75		1	1	4		
5.5		4	2	2		
5.25	2	3	1	1		
5.0			2	3		
4.75	2	7	6	3	1	1
4.5	5	4	4	5		1
4.25	6	6	4	11	3	1
4.0	5	3	4	3	2	1
3.75	3	4	7	4	2	2
3.5	6	3	6	2	2	4
3.25	4	1	4	2	3	1
3.0	4	2	1	3		
Less than 3.0	13	8	6	3		5
N	49	50	50	50	14	16

Calculation of correlation coefficients between *mental ability scores and motor skills scores* seemed unwise because three different tests had been used as mental measures, because these tests had not all been given by the same tester, and because the tests themselves are all somewhat inadequate as measures of the mental ability needed for success in industry. The totally blind group was therefore roughly separated into three sections: superior, having I.Q.'s of 110 or better; average, having I.Q.'s between 90 and 110; and inferior, having I.Q.'s below 90.

Distributions for each mental ability group on the six tests were prepared.[4] Mental superiority seems in general to have some advantage, mental inferiority some disadvantage. However, there is little difference in spread of scores between the average and superior groups. On Disassembly, Displacing, and Toolsample No. 1, the highest scores were

[4] All distributions mentioned in this paper are on file with the Library of the American Foundation for the Blind, New York, and can be borrowed by application to them.

Table 4

Distribution of Standard Scores of Men and Women with Better than 10/200 but not More than 20/200 Vision (Group C)

Standard Score	Displacing	Turning	Assembly	Disassembly	Tool-sample, Part 1	Tool-sample, Part 2
6.75		1		2		
6.5		1	2	1	1	
6.25		1		2	1	
6.0	2	4	2	1	1	
5.75		1	3	4		
5.5	3	3	3	5	1	
5.25	1	2	3	5	1	1
5.0	2	3	4	4	2	1
4.75	2	5	3		1	1
4.5	4	1	1	1	2	3
4.25	3	3	2	2	1	
4.0	3	1	2	4	3	
3.75	4	3	2		1	3
3.5	1	1	5			2
3.25	3	2			1	1
3.0				1		1
Less than 3.0	5	2	2	2		3
N	33	34	34	34	16	16

obtained by persons in the average group. Also, the best score for Turning for the mentally inferior group exceeds the best score for Turning for the mentally average group, while the best scores for Assembly for both groups are at the same point. Here again, as in the case of age, generalizations on the basis of mental ability may be made for the group but could be most unjustly applied to a given individual.

To study the significance of *age of onset of blindness* the group was subdivided into five parts: those blind at birth or before reaching one year of age, those who became blind at ages between one and nine years, those who became blind between 10 and 19 years, those who became blind between 20 and 29 years, and those who lost their vision after 30. Since our clients varied in age from 15 to 39 years, this type of subdivision did not indicate the length of time each group had been blind. A second division was therefore made, comparing the test results of persons blind more than ten years with those of persons blind less than two years.

On Displacing, Turning, Assembly and Disassembly, there is a slight but persistent advantage for those blinded early in life. Except on Turning, this advantage is even clearer when those blinded more than 10 years are compared with those blinded less than two years. Not only do a greater proportion of those blinded early in life or those blinded more than ten years before testing attain acceptable scores, but most of the really high scores are obtained by these groups. Both quantitatively and qualitatively a longer period of blindness results in a distinct advantage for the group.

With the two parts of Toolsample the reverse is true. Despite the preceding evidence that early blindness or prolonged blindness leads to improved orientation and general adjustment to manual tasks, a distinct superiority on Toolsample is shown by those groups which lost their vision in their teens or later.

A tabulation of results grouped by *cause of blindness* indicated no significant differences. Apparently a client is likely to do equally well on these tests whether the diagnosis of his blindness is Retinitis Pigmentosa, trauma, or any other cause. This may be partly explained by the fact that the causes of blindness are sometimes obscure, and in many cases more than one factor plays a part.

Study of Employed Group

In order to establish roughly some standard which would enable us to make use of our test results in recommending placements, before the completion of this study, nineteen blind persons who were successfully employed in industry were asked to take the tests.

As with persons having vision who have been successful in industrial employment, the members of this employed blind group rarely obtain standard scores below 4.0. In no case did one of them obtain more than one standard score below 4.0 out of five tests. The Displacing Test which subsequently grew out of our experience in testing blind persons, had not been developed at the time this group was tested. In no case did a client in this group deviate by more than one and one-half standard score units from the mean of the seeing population. As far as the test results are concerned, this, then became our criterion—that for prediction of successful employment in industry, the blind individual should be able to attain on these tests standard scores of 4.0 or better, or should fall below that level only slightly and only on one test.

Study of the Learning Curve

To determine the curve of learning of blind persons in repetitive manipulations, 51 blind students were given a period of training with the

Turning part of the Minnesota Rate of Manipulation and the two parts of the Pennsylvania Bi-Manual Worksample. Of these students, 27 were girls, 24 boys. They ranged in age from 14 years 7 months to 27 years 6 months, with an average age of 18 years. This period of training consisted of the repetition of the three tests, three times without rest between, on each of four successive days. The training was done under controlled conditions, where no interruptions or distractions were permitted. The establishment of equal motivation was, unfortunately, more difficult. It was not possible to work out any material reward which would be acceptable under the circumstances. The potential importance of this experiment was explained to the students but it is difficult to determine the extent to which this completely abstract motivation operated with students of varying degrees of maturity, varying intellectual capacities and varying backgrounds. In many cases great interest was expressed but in a few instances the reverse was true.

Despite this unfavorable circumstance with regard to motivation, the results approximate the familiar outlines of curves of learning found in many other studies. Ten students from a public high school were given the same training under similar conditions. Their results follow very closely the curves of the blind students.

Intercorrelations of Test Results

Since the test results for blind clients under 40 years of age do not fall into a normal distribution, the rank order method was used in obtaining the intercorrelations in Table 5. Further, because of certain changes made in methods of testing, only about half of the sets of results could be used to obtain these correlations.

All the correlations, excluding Toolsample, parts 1 and 2, are higher than those found in our studies of seeing clients at the Center. No statistical evidence is available at this writing to explain this but our

Table 5

Intercorrelations of the Test Results of Totally Blind Persons on the Six Manipulative Tasks

	Turning	Assembly	Dis-assembly	Tool-sample, Part 2	Tool-sample, Part 1
Displacing	.70 (70)	.72 (70)	.74 (70)	−.08 (32)	−.42 (32)
Turning		.58 (71)	.48 (71)	−.04 (32)	−.48 (32)
Assembly			.74 (71)	.27 (55)	−.10 (55)
Disassembly				.35 (55)	.02 (55)
Toolsample, part 2					.62 (55)

Note: The number upon which each coefficient is based is shown in parentheses following it.

observation suggests that the important part played by orientation in the work space increases the coefficient of correlation. This is especially evident in the correlations of the Displacing Test, in which it has already been noted in other connections that orientation plays a vital part. The negative correlations between Displacing and Turning versus Toolsample are probably further evidence of the relationship mentioned under the discussion of test results related to length of blindness; in our heterogeneous group, a number of the persons who were most at a disadvantage with the first two tasks had the advantage of some familiarity with the materials of the Toolsample.

Summary

1. Among persons who are legally classed as blind, the group which retains some degree of vision has a distinct advantage in the performance of concrete tasks, chiefly because of the ease with which they orient themselves in the work space. This generalization with regard to the group, however, does not necessarily apply to the individual. Many totally blind persons surpass others having close to 20/200 vision, and the amount by which they surpass them varies from task to task. Indeed, those blind persons who obtained standard scores above 5.00 surpassed at least fifty per cent of the seeing population upon which the norms were established.

2. Being more than 40 years of age may also be regarded as disadvantageous if group results only are considered, but in certain individual cases no disadvantage is found.

3. Mental inferiority is a disadvantage but it is not yet possible to set a definite level below which no individual should be considered for job placement and above which most will be successful. This situation may be mainly the result of the inexactness of the measures of mental ability which have been used with the adult blind.

4. Studies of test results in relation to age of onset of blindness suggest that orientation in space comes slowly. They also point very strongly to the desirability of including in the education of blind children far more contact with practical materials of possible vocational value, such as tools. The inability of some blind individuals to deal with this type of material suggests lack of previous contact with it rather than lack of fundamental ability to comprehend it.

5. Our studies give no reason for basing job placement on the cause of blindness as reported by the individual.

6. The curve of learning for blind persons takes the same general course as does that of seeing persons. The blind, however, require a greater number of repetitions before reaching their maximum performance. As a group, the blind will approximate the performance of the

seeing most nearly in those motor skills in which orientation in space is of least importance, and will remain farthest from the standard of the seeing where orientation in space is most important. Individual members of the blind group may, however, give an excellent performance even where orientation in space is a major factor.

7. In general, a higher correlation must be expected between tests of motor skills used with blind persons than between these same tests when used with seeing persons. This probably is due to the fact that blind individuals with superior orientation in space have an advantage on all motor skills tasks.

8. In testing and experimentation upon motor skills, motivation is more difficult to establish with the blind than with the seeing. The blind as a group live in a protected world where speed is relatively unimportant. In some cases, the very concept of speed seems foreign to them.

9. Tests can be of great value in the selection of individuals for job placement and in the selection of the job in which a chosen individual shall be placed. Their predictive value is by no means perfect, but that is also true of their predictive value when used with seeing groups. A more comprehensive group of motor skills tests is needed. A better measure of general mental ability, particularly sensitive to comprehension rather than mere verbalization, is needed. And, perhaps most of all, a reasonably reliable measure of emotional and personal adjustment is extremely needed.

10. Two points require final emphasis. The first is that vocational testing in general, and testing of the blind in particular, is still so inexact that it can be safely used only by a person with sound psychological training and good clinical experience. The importance of the qualitative aspects of the testing was mentioned earlier. This is a factor which cannot be subjected to statistical treatment or readily reduced to a written statement. Unless the test administrator has, from his clinical background, learned to make reliable observations and judgments of people, and unless tests have a meaning for him which goes beyond the bare words of an instruction manual, he should not jeopardize the future of a blind person by a rash diagnosis.

11. The second point is also an amplification of one that is true for all vocational guidance. The individual client is the only point of consideration. Studies like the present one lead to generalizations about the group but guidance and placement work deals with individual men and women. Variations of individuals from the average of the group are tremendous. Not tests alone, but medical findings and history, social, economic, and educational background, personality factors, and indications of present motivation, are vitally important considerations.

Received February 7, 1945.

The Significance of Verbal Aptitude in the Type of Occupation Pursued by Illiterates *

Capt. William D. Altus and T/Sgt. Clarence A. Mahler

*Ninth Service Command Special Training Center, Camp McQuaide, California***

The discharged veteran is accorded, in most cases, the right to further his education for one to four years, if he so desires, with government assistance. Proper vocational and educational counseling of soldiers returning to civilian life is highly important. One type of soldier who especially needs expert guidance if he desires vocational or other training is the soldier who was classified as an illiterate when he was inducted into the Army. If this man is to be tested as part of the guidance procedure, account must be taken of his reading handicap.

The universally employed group test may have adequate validity for those who are normally literate, but for those whose literacy is marginal the group test score is not always dependable as a criterion of aptitude. All soldiers, for instance, take the Army General Classification Test. In one study of soldiers processed at this Center for Army illiterates, there was an average increase of approximately twenty points in standard score on the AGCT from their first testing, prior to entry into the Center, until they were tested as graduates some weeks later. It is reasonable to assume that if these men had had somewhat more training in this or a similar center, where twelve weeks was the maximum time they could stay, they would show an even higher average increase in AGCT score.

In order to obviate some of the gross effects of illiteracy, as reflected in an aptitude score on a group test, the Personnel Consultants' Section in the Ninth Service Command Special Training Center used the Wechsler Mental Ability Scale, Form B, an individually administered test, as the criterion of verbal aptitude. This test was supplanted by the Army Individual Test in late 1944. Not all the subtests of the Wechsler were given; in fact, if the trainee was English speaking, he was given only four of the verbal subtests, Information, Arithmetic, Comprehension and Similarities. In this shortened form the scale proved to possess adequate validity in a battery of tests for initially selecting those men who

* The opinions expressed in this article are the authors' and do not necessarily reflect the official attitude of the Army of the United States.

** Center inactivated in November, 1945.

could not qualify as literate within the twelve weeks' time limit. The coefficient of biserial correlation of the Wechsler with the disposition of the trainee (whether graduated or discharged as inapt) is .521 for over four thousand cases. It is therefore clear that this individual test discriminated among the abilities of trainees so far as their disposition is concerned, even though all of these soldiers upon arrival were technically illiterate as the Army defines illiteracy.

It should be mentioned that, although these trainees are initially classified as illiterate and are therefore presumably homogeneous in that none reads very efficiently, they are not so homogeneous in verbal aptitude (or intelligence as it is commonly called). On the contrary, this Center may be said to have received approximately the lowest three sigmas, as measured on the Wechsler test, if the whole range of verbal aptitude among those who are self-supporting is divided into six sigma units. These men showed considerable variation in mental ability, much more than might be inferred from their classification as technical illiterates. Among some of the most capable of all these men were a few who were so illiterate that they could recognize few, if any, printed letters. Even if they were eventually graduated and shipped, a group aptitude test did these latter men a rank injustice.

In addition to pointing out the inadequacy of group aptitude tests for some of these technical illiterates, it is the purpose of this paper to present data concerning the validity of the Wechsler standard scores in terms of the occupations these trainees have pursued. These data can fortunately be obtained from the personnel card which was filled out for each trainee upon his arrival. Many individually administered tests, besides the Wechsler, were given the trainee; also, certain biographical data were obtained and recorded on the personnel card. Among the biographical data were included the occupations at which the trainee had been employed prior to his induction. His main job, the one he followed longest during his work history, is the one used in the present study. Frequently, his job immediately prior to induction was a short-term one in some war industry, one to which he obviously would not return as a peacetime occupation and one for which, in many instances, he would not have been hired if labor had not been extremely scarce. To be classified as a main occupation, a job must have been pursued for a minimum of one year, even for the youngest of trainees, though usually the amount of time spent on the job was at least three to five years. Only Negroes and old-line, native-born Whites were used in the tabulation in order to avoid the effect of language handicap among certain bilinguals as reflected in their tested ability. The personnel cards of 1,674 White and 790 Negro trainees were reviewed for this study.

There were altogether 142 different job categories for the two racial

groups. The Whites had held 114 different jobs; the Negroes, 78. The list of these jobs is too long to present here but the sixteen main occupations, which account for 81.07% of all the trainees studied are presented in Table 1. The larger list of total occupations included horse trader, entertainer, casket upholsterer, pinsetter, boxer, tree surgeon, motion picture projectionist, sheepherder, fur skinner, florist, dental technician, barrel maker, bookbinder and so on.

It will be seen in Table 1 that agricultural pursuits claimed over half of the 1,998 trainees here presented. There are significantly more Whites in the agricultural group than Negroes; the critical ratio of the difference in percentages is well over the conventional three. Though there were more farmers among the Whites, there were more Negroes who had engaged in truck driving. The Negroes also show a higher proportion of general and construction laborers. More Whites had been engaged as tractor drivers but there was a greater incidence, percentagewise, of cooks and mechanics' helpers among the Negroes. One occupation, that of porter, appeared to operate solely through color selection, for while 33 Negroes had been so engaged none of the Whites appeared in this category. For the other occupations listed in Table 1, the group differences in percentage are not large.

Table 1

The Sixteen Main Occupations of White and Negro Trainees

Occupation	White		Negro		Total	
	N	% of all White	*N*	% of all Negro	*N*	% of Both Negro and White
Farm Hand	856	60.75	194	32.94	1050	52.56
Truck Driver	187	13.27	115	19.52	302	15.13
Laborer, Gen'l	85	6.03	73	12.39	158	7.91
Laborer, Const.	48	3.41	33	5.60	81	4.05
Cook	18	1.28	32	5.44	50	2.50
Mechanic's Helper	28	1.99	22	3.74	50	2.50
Tractor Driver	34	2.41	5	.84	39	1.95
Miner	34	2.41	4	.68	38	1.90
Sawmill Worker	14	.99	22	3.74	36	1.80
Porter	0	.00	33	5.60	33	1.65
Section Hand, R. R.	20	1.42	12	2.04	32	1.60
Service Station Attendant	12	.85	20	3.39	32	1.60
Painter	18	1.28	11	1.87	29	1.45
Welder	20	1.42	7	1.19	27	1.35
Lumberjack	19	1.35	2	.34	21	1.05
Carpenter	16	1.14	4	.68	20	1.00
Totals	1409	100.00	589	100.00	1998	100.00

When the whole group of 2,476 trainees was classified by major occupational groups and divisions of the *Dictionary of Occupational Titles*, it was found that one in eleven had held a skilled job, such as carpenter, baker, miner, auto mechanic, painter, plumber, tailor and the like. One in four held a semi-skilled job, such as tire capper, roofer, presser, machinist's helper, lumberjack, truck driver or carpenter's helper. Roughly one in seven held an unskilled job. The other four broad groupings included the agricultural, fishery, forestry and kindred occupations (one broad category), the service occupations, clerical and sales and, lastly, the professional and managerial. The fact that there were 25 trainees in this last category arises from the peculiarities of the classification system: Boxers, dancers and entertainers are classified with the professional group. Two belonged legitimately to the managerial class; one of them owned and ran a cafe and another managed a lumber yard. Only six were classified with the clerical and sales group while 154 were in the service occupations. About 40% of the total fitted into the agricultural, fishery, forestry group.

Those engaged in agriculture were classified with the unskilled for the brief study which follows. The average standard score on the four verbal subtests of the Wechsler was then computed for the skilled, the semi-skilled and the unskilled, the Whites and Negroes being separately tabulated. When the tabulation was completed, it was found that the skilled and semi-skilled workers among both the Whites and Negroes had a Wechsler mean score which was significantly higher than that of the unskilled. There were also 87 chances in 100 that the skilled Negro worker's mean aptitude score was reliably higher than that of the Negro semi-skilled worker; for a similar comparison among the White groups, there were 96 chances in 100 that skilled worker's higher mean score was not due to chance factors. Even though all are technically illiterate when they arrive at the Center, it is clear that the skilled and semi-skilled workers, whether White or Negro, are reliably brighter than the unskilled if the Wechsler mean score is equated with brightness. The Wechsler also comes close to differentiating the skilled worker from the semi-skilled in terms of verbal aptitude.

The second study was addressed to the extremes in tested aptitude. The brightest ten per cent, among both White and Negro groups, earn approximately 28 or more points in standard score on the four verbal subtests of the Wechsler while the lowest ten per cent earn from zero to twelve points. There were 100 men in each of the four categories, that is, there were 100 White and 100 Negro trainees scoring from zero to twelve points and a like number scoring 28 or more points. The categorizing as skilled, semi-skilled and unskilled was accomplished, wherever possible, by the use of the *Dictionary of Occupational Titles*. Where the

men did not fit into one of three categories of skill but fitted into one of the other broad occupational groupings, they were shifted into one of the three divisions of relative skill by a procedure that could be called rule-of-thumb.

Table 2 shows that three times as many Whites and almost twice as many Negroes scoring 28 or more points on the Wechsler were classified as skilled as was true of the comparable low-scoring group. Almost three times as many of the higher-scoring Whites were classified as semi-skilled as was found among the low-scoring group; the differential was even more marked for the Negroes. Almost four-fifths of the low-scoring

Table 2

Occupational Skills and Extremes in Wechsler Aptitude

	Skilled		Semi-Skilled		Unskilled	
Group	Bright	Dull	Bright	Dull	Bright	Dull
White	18	6	45	17	37	77
Negro	9	5	49	15	42	80

White and Negro groups are to be found among the unskilled, roughly twice the figure found for the high-scoring trainees. Aptitude, even among the technically illiterate, varies significantly according to the type of occupation pursued.

Not shown in Table 2 is the number of occupations engaged in by these groups of varying aptitude. The 100 relatively more capable Negroes had been employed in 39 different occupations; the 100 relatively duller Negroes had engaged in only 30 occupations. For the White groups the figures are even more marked: The high-scoring group had engaged in a total of 37 different occupations, while the low-scoring group had engaged in only 18. This finding is consistent with what would logically be expected, that is, that the brighter the individual the greater the range of occupational choice open to him. Verbal aptitude thus appears to be associated with not only the type of occupation followed by these technically illiterate men of limited mentality but also with their range of occupational choice.

One can never be certain of interpreting correctly a low score on a group test of aptitude when the score is earned by a marginally literate individual. Literacy and verbal aptitude are positively correlated, to be sure, even in such a restricted group as is found in an Army Special Training Center. Nevertheless, the correspondence is not sufficiently high for one to assume that all illiterates are automatically dull. As has been previously said, some of the most verbally apt trainees in Wechsler score have arrived at this Center total illiterates.

Only eight to ten minutes are required to administer the shortened form of the Army Wechsler here employed. Though taking little administration time, the four subtests have been found to discriminate quite satisfactorily among the varying learning capacities of the trainees, so far as the acquisition of literacy is concerned, and also to be associated with occupational levels, the brighter in score tending much more often to have followed a more highly skilled job. For these reasons, it is recommended by the authors that those who counsel these men when they return to civilian life use an individually administered test of comparable validity as a necessary part of the evaluative process, particularly if the counselling is concerned with occupational choice or occupational training.

Received February 3, 1945.

Readability of Newspaper Headlines Printed in Capitals and in Lower Case

Donald G. Paterson and Miles A. Tinker

University of Minnesota

The question as to the relative legibility of headlines set in all capital letters and of headlines set in lower case has long been in dispute among newspaper men. In the absence of scientific studies of the problem it is obvious that the question must remain one for debate.

In the course of our investigations of typographical factors influencing speed of reading we demonstrated that paragraphs printed in all capitals are read 11.8 per cent more slowly than paragraphs printed in caps and lower case. Few typographical factors exert such a retarding effect on reading speed.

In presenting these results in *How to make type readable*,[1] the writers strongly recommended that all capitals printing be eliminated where speed of perception is at a premium. It was pointed out that printing in all capitals survives in those situations where speed of perception is especially important namely, posters, car cards, billboards, newspaper headlines, etc. The headline writer was urged to abandon all capitals and to rely primarily on lower case type.

Within a few months after those recommendations were published, a long and lively debate by correspondence between the writers and Theodore Bernstein of the *New York Times* was initiated by Bernstein who challenged our right to generalize from the normal reading of continuous text to the perception of short headlines.[2] He pointed out the essential differences between these two situations and argued that all capital headlines would be perceived more quickly and at a greater distance.

The debate was terminated by mutual agreement that the question should be settled by direct experimental attack.

The writers persuaded two graduate students to undertake a first study of the question by means of the orthodox brief exposure or tachis-

[1] Paterson, D. G., and Tinker, M. A. *How to make type readable.* New York: Harper and Brothers, 1940, pp. 22–25. (Obtainable from the authors.)

[2] Curiously enough, no mention is made regarding the relative legibility of all capitals vs. lower case headlines in the excellent book by Garat, R. E., and Bernstein, T. M., entitled: *Headlines and deadlines.* New York: Columbia University Press, 1933.

toscopic technique. Five word single column headlines set in 24 pt. Cheltenham bold face were printed in all capitals and in caps and lower case and the relative legibility of the two kinds of headlines was determined at the normal reading distance of about 15 inches. Breland and Breland's findings [3] disclosed an 18.9 per cent difference in favor of lower case single-column headlines. English likewise found a similar difference.[4] In other words, lower case printing was found to be even more efficient in the single-column headline situation than in the continuous text situation.[5]

Realizing that Breland and Breland's findings might not hold for single-column heads when read at a distance, nor for multi-column or banner headlines which are intended to be read from afar, the writers persuaded another graduate student (Miss Alice Warren) to undertake a study of these additional problems. There was no assurance that a similar result would be found. Indeed, an opposite result might well have been expected in view of Tinker's previous demonstration that "both capital letters and words in capitals were read at greater distances from the subject than letters or words in lower case." [6]

The reader should keep in mind, however, that the greater perceptibility at a distance of capital letters and words in capitals was a function of perceptibility without reference to *speed* of perception. Furthermore, the experiment was confined to the perceptibility of isolated letters and isolated words and not to the legibility or readability of meaningful phrases such as are used in multi-column headlines. The headline problem is essentially one of the apprehension *at a glance* of the meaning of a group of related words. Under these conditions single-column and multi-column headlines in caps and lower case might be found to be apprehended at a glance even at a greater distance than when such headlines are set in all capitals.

Warren's first study utilized the same single column headlines as were used in the Breland and Breland study but instead of being exposed by means of the Dodge Tachistoscope at a distance of 15 inches from the

[3] Breland, K., and Breland, M. K. Legibility of newspaper headlines printed in capitals and lower case. *J. appl. Psychol.*, 1944, 28, 117–120.

[4] English reports a reading loss of 18 per cent for headlines set in regular all caps in a comparison with headlines set in caps and lower case. See E. English, A study of the readability of four newspaper headline types. *Journalism Quarterly*, 1944, 21, 217–229.

[5] In a letter to M. K. Breland, Bernstein wrote, "Even though I must confess that the results showed a more pronounced difference than I had expected, I am gratified that at least I had some part in advancing scientific backing for a hypothesis that had been airily affirmed for years without such backing. Your paper is important in the annals of both typography and journalism."

[6] Tinker, M. A. The influence of form of type on the perception of words. *J. appl. Psychol.*, 1932, 16, 167–174.

subject, they were exposed by means of the Whipple Disc Tachistoscope at a distance of five and one-half feet from the subject.[7] The exposure time in the Breland and Breland study was 50 milliseconds (1/20th of a second) whereas in the Warren study the exposure time was somewhat longer, i.e., 180 milliseconds (5/50ths of a second). Both studies tested one subject at a time.

The results of Warren's first study reveal that, under the conditions as defined, lower case single-column headlines and all capitals single-column headlines are equally legible. There was an infinitesimal difference of .3 of one per cent but from a statistical point of view and from a practical point of view both kinds of single-column headlines are equally legible at a distance of five and one-half feet. This is the distance one encounters in reading a headline over someone else's shoulder, or across a breakfast table.

In this part of the study, each subject, after completing the test, was asked whether he had noticed any differences in the appearance of the headlines and if so, which kind appeared to be easiest to read. Of the 40 subjects, 25 noted no difference, 8 believed lower case was easier to read and 7 believed upper case was easier to read. From the point of view of the headline reader, therefore, neither kind of headline appeared to have any advantage over the other.

The second study, conducted by Warren and reported in her thesis, was aimed at determining the relative legibility, at various distances, of 100 multi-column or banner headlines when set in lower case and in all capitals.

The headlines selected for study were taken directly from those actually used in Minneapolis papers between 1934 and 1940. They were printed directly on rectangular pieces of cardboard 28″ × 7¼″ from the largest type immediately available in the University Printing Shop, namely, 60 point size in Memphis bold face.

Since no tachistoscope suitable for exposing such large surfaces was available Miss Warren arranged a stand and exposed each banner headline for approximately one second by removing and replacing a large screen. The group method of testing was employed which permitted determination of the legibility of the two kinds of banner headlines at different fixed distances from approximately 6′2″ to 17′1″ and at different angles. The student sitting in the center of row one was 6′2″ from the exposure stand and those sitting at the end of row one were 7′11″. The student sitting in the center of row two was 9′10″ from the stand and those at the end of row two were 10′9″ away. In row three, the distances

[7] Warren, A. L. *The perceptibility of lower case and all capitals newspaper headlines.* Master's thesis, University of Minnesota Graduate School, November 1942. On file in University of Minnesota Library.

toscopic technique. Five word single column headlines set in 24 pt. Cheltenham bold face were printed in all capitals and in caps and lower case and the relative legibility of the two kinds of headlines was determined at the normal reading distance of about 15 inches. Breland and Breland's findings [3] disclosed an 18.9 per cent difference in favor of lower case single-column headlines. English likewise found a similar difference.[4] In other words, lower case printing was found to be even more efficient in the single-column headline situation than in the continuous text situation.[5]

Realizing that Breland and Breland's findings might not hold for single-column heads when read at a distance, nor for multi-column or banner headlines which are intended to be read from afar, the writers persuaded another graduate student (Miss Alice Warren) to undertake a study of these additional problems. There was no assurance that a similar result would be found. Indeed, an opposite result might well have been expected in view of Tinker's previous demonstration that "both capital letters and words in capitals were read at greater distances from the subject than letters or words in lower case." [6]

The reader should keep in mind, however, that the greater perceptibility at a distance of capital letters and words in capitals was a function of perceptibility without reference to *speed* of perception. Furthermore, the experiment was confined to the perceptibility of isolated letters and isolated words and not to the legibility or readability of meaningful phrases such as are used in multi-column headlines. The headline problem is essentially one of the apprehension *at a glance* of the meaning of a group of related words. Under these conditions single-column and multi-column headlines in caps and lower case might be found to be apprehended at a glance even at a greater distance than when such headlines are set in all capitals.

Warren's first study utilized the same single column headlines as were used in the Breland and Breland study but instead of being exposed by means of the Dodge Tachistoscope at a distance of 15 inches from the

[3] Breland, K., and Breland, M. K. Legibility of newspaper headlines printed in capitals and lower case. *J. appl. Psychol.*, 1944, 28, 117–120.

[4] English reports a reading loss of 18 per cent for headlines set in regular all caps in a comparison with headlines set in caps and lower case. See E. English, A study of the readability of four newspaper headline types. *Journalism Quarterly*, 1944, 21, 217–229.

[5] In a letter to M. K. Breland, Bernstein wrote, "Even though I must confess that the results showed a more pronounced difference than I had expected, I am gratified that at least I had some part in advancing scientific backing for a hypothesis that had been airily affirmed for years without such backing. Your paper is important in the annals of both typography and journalism."

[6] Tinker, M. A. The influence of form of type on the perception of words. *J. appl. Psychol.*, 1932, 16, 167–174.

subject, they were exposed by means of the Whipple Disc Tachistoscope at a distance of five and one-half feet from the subject.[7] The exposure time in the Breland and Breland study was 50 milliseconds (1/20th of a second) whereas in the Warren study the exposure time was somewhat longer, i.e., 180 milliseconds (9/50ths of a second). Both studies tested one subject at a time.

The results of Warren's first study reveal that, under the conditions as defined, lower case single-column headlines and all capitals single-column headlines are equally legible. There was an infinitesimal difference of .3 of one per cent but from a statistical point of view and from a practical point of view both kinds of single-column headlines are equally legible at a distance of five and one-half feet. This is the distance one encounters in reading a headline over someone else's shoulder, or across a breakfast table.

In this part of the study, each subject, after completing the test, was asked whether he had noticed any differences in the appearance of the headlines and if so, which kind appeared to be easiest to read. Of the 40 subjects, 25 noted no difference, 8 believed lower case was easier to read and 7 believed upper case was easier to read. From the point of view of the headline reader, therefore, neither kind of headline appeared to have any advantage over the other.

The second study, conducted by Warren and reported in her thesis, was aimed at determining the relative legibility, at various distances, of 100 multi-column or banner headlines when set in lower case and in all capitals.

The headlines selected for study were taken directly from those actually used in Minneapolis papers between 1934 and 1940. They were printed directly on rectangular pieces of cardboard 28″ × 7¼″ from the largest type immediately available in the University Printing Shop, namely, 60 point size in Memphis bold face.

Since no tachistoscope suitable for exposing such large surfaces was available Miss Warren arranged a stand and exposed each banner headline for approximately one second by removing and replacing a large screen. The group method of testing was employed which permitted determination of the legibility of the two kinds of banner headlines at different fixed distances from approximately 6′2″ to 17′1″ and at different angles. The student sitting in the center of row one was 6′2″ from the exposure stand and those sitting at the end of row one were 7′11″. The student sitting in the center of row two was 9′10″ from the stand and those at the end of row two were 10′9″ away. In row three, the distances

[7] Warren, A. L. *The perceptibility of lower case and all capitals newspaper headlines.* Master's thesis, University of Minnesota Graduate School, November 1942. On file in University of Minnesota Library.

were 13′4″ in the center and 13′8″ at the end. In row four the distances were 16′10″ and 17′1″. Thus, those sitting in row one were glimpsing the headlines at a distance approximating the situation where one looks at a headline across a streetcar, or on a rapidly passed newsstand. On the other hand those sitting in rows two to four were at distances much greater than would ordinarily prevail in daily life. That is, newspaper publishers do not expect that their banner headlines will be read at distances much beyond seven feet. The results, therefore, for row one will be especially significant and should approximate the findings for the first part of Miss Warren's study where each subject was 5½ feet from the tachistoscope.

The results are presented in Table 1. At the first row distance (6′2″ to 7′11″) the lower case banner headlines seem to be superior to those set in upper case. The loss in perceptibility was 5.3 per cent when the headlines were set in all capitals. This lower case superiority was present for practically all of the students sitting in row one (10 out of 12 made better scores in reading the lower case headlines). The difference is statistically significant at between the 2 per cent and 5 per cent levels of significance.[8]

Table 1

The relative perceptibility, at different distances, of five-word 60 point banner headlines set in lower case and in all capitals

Note: Number of lower case headlines = 50; number of upper case headlines = 50. Perfect score would be 250 words correctly apprehended in lower case headlines (5 × 50) and 250 for upper case headlines. Total number of subjects (N) = 46.

Row	Distance from Row to Exposure Stand: Center	Distance from Row to Exposure Stand: End of Row	N	Type	Mean Number of Words Perceived	Difference in Per Cent
1	6′2″	7′11″	12	Lower Case	132.3	− 5.3*
				Upper Case	125.3	
2	9′10″	10′9″	12	Lower Case	117.5	+ 2.6
				Upper Case	120.5	
3	13′4″	13′8″	10	Lower Case	102.4	+ 4.8
				Upper Case	107.3	
4	16′10″	17′1″	12	Lower Case	79.4	+19.8*
				Upper Case	95.2	

* Statistically significant.

[8] This is a technical statistical interpretation based on Fisher's t test of significance and indicates a high degree of probability that the obtained difference would be found to be in the same direction in from 95 to 98 times of 100 repetitions of the experiment.

At first thought one may be inclined to believe that the results of the several studies are inconsistent. The Breland and Breland study revealed a striking superiority of single-column lower case headlines over upper case headlines when perceived at the *normal reading* distance of about 15 inches. The first part of the Warren study showed that both kinds of single-column headlines were equally effective when glimpsed at a distance of 5½ feet. But one should keep in mind that single-column headlines are not intended to be read at any such distance. The second part of Warren's study again discloses the superiority of lower case banner headlines when read at a distance of from six to seven feet which is about the distance at which banner headlines are supposed to be read. Thus, in those situations in which single column or in which multi-column headlines are supposed to be read, it is clear that lower case headlines are distinctly superior to upper case headlines.

The results for rows two and three as given in Table 1 indicate a slight advantage for upper case headlines but the difference is not statistically significant. Furthermore, it is important to note that the average scores for those sitting in rows two and there are lower than the mean score for those sitting in row one. This means that those stationed beyond seven feet from the exposure stand had increasing difficulty in correctly apprehending the headlines even though the exposure time was relatively long (one second).

When we inspect the results for row four in Table 1 it is apparent that the correct apprehension of the five word headlines becomes quite difficult and upper case banner headlines have a distinct advantage. The reading of 60 point banner headlines at approximately 17 feet is thus quite difficult and inaccurate. Apparently the subjects were forced to pick out a word here and there during the one second exposure period. In this sort of situation upper case headlines come into their own since they are larger and the letter outlines are more distinct. This result is quite in line with the earlier study of Tinker.[9]

Our interpretation of the results of the experimental studies now available would lead us to emphasize, as we did originally, the importance of *characteristic word form* which is provided by lower case printing in promoting readability.[10]

At least three reasons can be found which will account for the poor legibility of all capitals.

[9] M. A. Tinker, *op. cit.*

[10] Eye-movement photography in reading all capitals text and in reading lower case shows a 12.4 per cent *increase* in fixation frequency with fewer words perceived for the reading of all capitals. See Tinker, M. A., and Paterson, D. G. Influence of type form on eye movements. *J. exp. Psychol.*, 1939, 25, 528–531.

(1) The printing surface required for upper case headlines is much greater than for the same headlines set in lower case *in the same point size.* A 60 point banner headline is 37.5 per cent longer when set in upper case than when set in lower case. Thus, each visual fixation permits the apprehension of a larger amount of meaningful material in each portion of a headline set in lower case than when the same headline is set in upper case.

(2) The word form is far more characteristic when words are printed in lower case than when they are printed in all capitals. Figure 1 shows that words printed in upper case have a rectangular

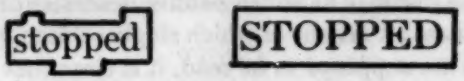


FIG. 1. Block outlines of the printed word "stopped" to illustrate that lower case exhibits a characteristic "word form" whereas "word form" is absent when printed in all capitals.

outline, one word differing from another merely in length whereas words printed in lower case or in caps and lower case have an irregular block outline, one word differing from another in a distinctive fashion. This fosters the reading of words as wholes. Figure 2 shows that the

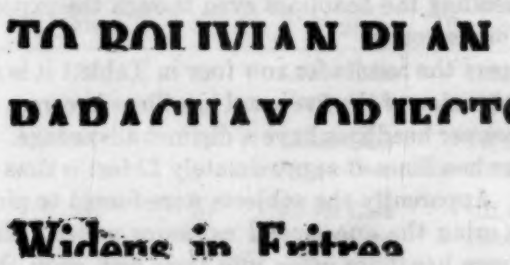


FIG. 2. Showing that the upper half of a printed line furnishes more cues to "word form" when printed in lower case.

upper half of a printed line furnishes more cues to "word form" when set in lower case than when set in upper case with the result that its meaning is more readily apprehended.

(3) Reading habits favor lower case since practically all of our reading is devoted to the apprehension of material set in lower case rather than in upper case.

It seems perfectly clear that our original recommendation that upper case printing be abandoned except for an occasional use when one desires

to attract attention through novelty is fully justified. One may now safely conclude that even in billboard printing where reading is to be done at great distances it would be preferable to rely upon caps and lower case than on all capitals. This may seem illogical to some who might be inclined to cite the results reported here as justifying the use of upper case printing for materials to be read at distances beyond seven feet. But it should be pointed out that in all these comparisons we have been observing the effect of using upper and lower case printing *in the same type size.* If one wants banner headlines to be read at a distance greater than seven feet then one need only increase the size of the type from 60 to 72 point or larger and by specifying the use of lower case all of the demonstrated advantages of characteristic word forms as cues to speedy and accurate perception will be retained. One should not forget that larger sizes of lower case than of upper case can be used within a given space. For example, a 60 point banner headline 10″ in length in lower case is 13¾″ long in upper case. This principle can be extended to car cards, poster signs and billboards which are intended to be read at far greater distances than is true of newspaper headlines. As a matter of fact, we can not think of a single type of reading situation in which upper case printing would be advantageous. A possible exception might be where the perception of an isolated letter to be used as a signal were required as along a railroad or highway. In such rare situations, an upper case letter might be preferred.

Summary

1. The problem of the relative legibility of headlines set in all capital letters and of headlines set in lower case has long been a subject of debate among newspaper men. Scientific studies bearing directly on the problem have been lacking.

2. Several years ago the writers proved that continuous text set in all capitals was read about 12 per cent more slowly than text set in lower case. Generalizing from these results to headlines, the writers recommended the abandonment of all capitals printing. This recommendation was sharply challenged by T. M. Bernstein of the *New Tork Yimes* so the writers arranged to have three investigations conducted that would bear directly on the problem at issue.

3. The first study dealt with five word single-column headlines set in 24 point bold face, half in upper case and half in lower case to be read by tachistoscopic exposure at the normal reading distance of 15 inches. The results disclosed an 18.9 per cent difference in favor of the lower case headlines.

4. The second study used the same single-column headlines but exposed them at a distance of five and one-half feet. This time both kinds of headlines were shown to be equally legible.

5. The third study used multi-column or banner headlines set in 60 point bold face which were exposed at varying distances from six feet to seventeen feet. At a distance of six feet the legibility of the lower case banner headline was 5.3 per cent greater than the legibility of those set in upper case. At distances from ten feet to fourteen feet, both kinds of headlines were equally legible. At a distance of seventeen feet, however, the upper case headlines proved to be more legible.

6. The writers conclude that the findings fully justify their earlier recommendation that headline writers should abandon the use of upper case headlines. Caps and lower case printing provides *characteristic word forms* which serve as cues to the rapid and accurate reading of meaningful material. This advantage has now been shown to hold not only for the reading of continuous text but also for the apprehension at a glance of five word single-column and of five word multi-column headlines when exposed at the distances at which they are ordinarily supposed to be read. For reading at unusually great distances such as in billboard display advertising, the writers would insist that lower case is also to be preferred. In this special situation, a larger point size of lower case than of upper case would have to be specified.

7. In conclusion, the writers desire to repeat their earlier recommendation, namely: THE HEADLINE WRITER AND THE ADVERTISING COPY WRITER SHOULD HENCEFORTH ABANDON ALL CAPITALS and should rely on lower case. When this rule is generally adopted, an occasional use of all capitals might be justified as a device for attracting attention.

Received March 2, 1945.

Explorations in Personality by the Sentence Completion Method

Amanda R. Rohde

National Hospital for Speech Disorders, New York City

Clinical psychologists, vocational guidance counselors, teachers, and other professional persons who deal with youth problems have continuing need to become intimately acquainted with the needs, inner conflicts, fantasies, sentiments, attitudes, aspirations, and adjustment difficulties of the individuals requiring their services.

Most workers depend largely on the interview and the study of guidance records to discover the intimate traits and tendencies of those whom they counsel. The worker often prefers to supplement the interview with questionnaires, inventories, and personality tests to gain a more complete picture of the individual's traits. Recent years have witnessed a rapid expansion in this field.

All experienced workers realize that the direct question of the interview, the items to be answered or checked in the questionnaires, and inventories have distinct limitations for personality exploration of the deeper sort. Direct questioning tends to make the individual self conscious; it puts him on the defensive, usually preventing him from disclosing his true self. Freedom of expression is eliminated; the answers are controlled by the questions, the questions in fact suggest the answers.

Fortunately, the so-called projection methods for personality study avoid the resistance that is often met in direct questioning regarding personal matters. Diagnosis by projection techniques reveal latent needs, sentiments, feelings, and attitudes which the subject would be unwilling or unable to recognize or to express in direct communication.

The experimental work with projection techniques has been reviewed by Frank (1, 2) who originated the term, by Murray (6, 7), Horowitz and Murphy (3), Symonds and Samuels (10) and Symonds and Krugman (11).

Projective techniques include: free word association, interpretation of ink blots (Rorschach), picture interpretation (the Murray Thematic Apperception Test, designed to reveal fantasy life; and Stern's Cloud Pictures), autobiography, story completion, story writing, story telling, play techniques with young children, puppet shows, the psycho-drama, music reverie, and expression through the graphic and plastic arts.[1]

[1] Descriptions of tests mentioned may be found in the publication by Gertrude Hildreth, *Bibliography of mental tests and rating scales.* New York City: Psychological Corporation, 1939.

The common element in all of these procedures is the diversion of the subject's attention from his own psychic processes and himself as the respondent to the task in hand. Thus he may be led to divulge deep-seated feelings and tendencies of which even he himself may be unaware, and without sensing that he is revealing personal data.

The free word association technique in which the subject is to respond to each stimulus word, e.g. "love" "hate," with the first word that comes to mind, has been widely used for theoretical and practical purposes in determining attitudes, interests, for detecting falsification, emotional states, and conflicts and complexes.

The Sentence Completion Technique

The sentence completion devise in which the subject is asked to read to himself the forepart of a sentence and to write anything he wishes to complete the sentence is essentially a projection technique utilizing free association. In unconstrained response to sentence beginnings, the subject inadvertently reveals his true self, since there is no way in which he can anticipate the significance of his answers for personality study.

In 1928 Payne (8) published a sentence completion test of fifty items that was widely used in vocational counseling. Payne devised the test for use with college students to reveal personal traits through eliciting inhibited responses.

Tendler (12) has developed a twenty-item sentence completion test in which each item except the first begins with "I," for example: I feel happy when_____.

Lorge and Thorndike (4, 5) experimented with a 240-item sentence completion test in which the subject was restricted to single-word responses. They reported negative results in using the device for personality study, since they found little relation between test responses and the actual behavior of the subjects tested. These results are not surprising since the subjects were restricted to single-word replies; and judging from the samples of items given in the report, the items were none too carefully worded for the intended purpose. The device appears to be better suited, as is the classical free word association method, to lie detection and to the discovery of specific objects about which mental complexes center. In a single-word response the subject's expression is too highly controlled for him to reveal much about himself no matter how many items are used.

Sanford *et al.* included a "Completion of Sentences Test" among other projection techniques in an extensive study of children (9). The test presented the subjects with a series of 30 incomplete sentences—clauses or phrases—so constructed as to permit a wide variety of gram-

matical conclusions. By selecting systematically the content of the stimuli it was possible to limit the range of responses provided the subjects followed the rules of grammar in their answers. The intention was to explore certain response areas. The following classes of stimuli were included. (1) A series of press. The given phrase or clause described an environmental situation or event affecting some actor and was so arranged that a grammatical completion gave an account of the actor's response, an action pattern which could usually be classified according to the scheme of needs. (2) The first part of the sentence described an action pattern of some need and was so arranged that to complete the sentence grammatically was to give the situation or event (press) which aroused the need. (3) The phrase or clause given was ambiguous in that the action was incomplete or lacking in motivation and so constructed that a grammatical completion supplied the missing motive (need).

The test was administered to children in grades 3 to 9 inclusive. The results were combined with those of a "Completions of Pictures Test" and an "Interpretations Test" since it was concluded that not any one of these by itself could add much to the knowledge of the general project. Thus, regarding quantitative results, each of the tests was considered to remain in an exploratory stage of development. In so far as qualitative results were analyzed, the answers revealed consistent patterns of response as well as responses of no significance. However, the difference of response—even of the same stimulus material—indicated fairly clear differences of personality among the children tested.

Revision and Extension of the Payne Sentence Completion Blank

From experience with the original Payne Sentence Completions Blank, the writer became convinced that the method merited wider use as a projective technique for studying the personal characteristics of young people. In 1939, having obtained permission from the author's heirs for this purpose, a revision and extension of the Payne completions blank was worked out. The aim was to develop an instrument which would have the advantages of other projective techniques, but simplicity of procedure and interpretation to give practical value for psychological diagnosis in schools and other institutions where the study of large numbers of individuals is required.

Considerable preliminary experimentation was required to select the proper range of items, to phrase the items, and to arrange the order of items within the test so that the minimum number of items would be included that would yield maximum indications for diagnosis.

The criteria used in selecting and constructing items were as follows: (1) The range of different stimuli must be broad enough to elicit informa-

tion concerning all phases of personality. (2) The responses must be controlled as little as possible by the stimulus phrases so that the subject may have freedom of expression. (3) The total time required for the test must not exceed a period convenient for schedules of schools and institutions.

Among the items finally retained were some that were freer of control than others. However, they were on the whole somewhat less controlled than Payne's original items. Of these, thirty-five were retained in original form, four in modified form; the remainder were dropped and twenty-five new ones added. The final test consists of sixty-four items.[2] The following are illustrative sentence beginnings used in the questionnaire:

My school work.....
I want to know.....
There are times.....
My greatest longing.....
Girls.....
Earning my living.....
At night.....
I cannot understand what makes me.....
My father.....
When I.....
Death.....
My eyes.....
Love.....
At home.....
I become embarrassed.....

Following the last item there is a blank space with the instructions: Write below anything that seems important to you.

The Rohde-Hildreth Sentence Completion Blank is arranged as a four-page folder. On the cover page, space is reserved for identification data and for a summary of the test findings.

A code system may be used if the examiner prefers to instruct examinees not to write their names on the blanks. When the test is administered to groups, more freedom of expression is usually obtained if this procedure is adopted. It is recommended that the questionnaire be used during a regular class period, preferably an English class. The examiner instructs the students to complete the sentences that are partially begun and informs them that any response they care to make to any item will be entirely acceptable. If the blank be used for adults in other situations, it usually suffices merely to say that it is a sentence completion exercise.

There is no time limit. The subject proceeds at his own pace. Some people finish in half an hour; others require an hour and a half. The median time is 45 to 50 minutes.

[2] Sentence Completions by Amanda R. Rohde and Gertrude Hildreth, copyright 1940.

The questionnaire is intended for use with individuals approximately twelve years of age and above, including adults. Its use as a group device does not invalidate the results for individual diagnosis.

A standard scoring scheme has been worked out by means of which the results may be expressed in quantitative form.[3]

While the subject's needs (motives), inner states (conflicts, dejection, optimism, etc.), traits, and the press (environmental forces) revealed are the chief bases of quantitative results, many other important aspects of personality are manifested. One gets a fairly good idea of his tastes and sentiments, something of his idealogy or philosophy, and the general representation of the ego structure, besides indications of intellectual status and emotional maturity.

Incidentally, the subject's language fluency, diction, grammar, ability to spell and punctuate, as well as the legibility and character of handwriting may be taken into consideration in evaluating the responses of the individual.

Compared with other projection devices, the sentence completion blank is time saving in two ways: First, it may be administered effectively to groups of persons at one time. Secondly, even if it be used individually, after the initial directions are given, the instructions make it self-administering.

Whether or not a subject responds "truthfully" is immaterial from the standpoint of exploring the individual's traits and tendencies. Repressed feelings, attitudes, and ideas emerge regardless of whether the subject be "telling the truth" or not. One subject commented: "I've answered this as though I were another person." This fact did not vitiate the results as one might be inclined to assume; on the contrary, this significant comment only added to the validity of the personality picture furnished by the test responses.

The fact that an individual student may suspect the purport of the test does not invalidate the technique. Attempts at deception, false or cryptic answers of suspicious persons, are not less revealing than those of others who give full, free responses; because each person always writes what to him seems a good response, thus projecting his personality regardless of his intentions.

The sentence completion device has proved to be a valuable adjunct to the interview and in the advisement connected with vocational guidance and orientation courses. It is especially recommended for the study of persons who present adjustment problems.

[3] A manual with directions for administration and scoring has been prepared by the writer. Copies of the test and manual may be obtained from the writer, 411 West 115th St., New York City, 25, N. Y.

Although the device has been used successfully by teachers and guidance workers in understanding individual adjustment problems, the most penetrating analyses of results are possible only in the hands of clinicians with extensive psychological training and experience.

Validation of the Rohde-Hildreth Sentence Completion Technique

Since 1939 the revised sentence completion test has been used experimentally by the authors and selected workers, chiefly with students of high school age. It has also been used to study the personal adjustments of patients in the National Hospital for Speech Disorders of New York City, of college students, of public health nurses, and it has been tried out experimentally by psychologists in one of the U. S. Navy Hospitals.

In appraising the Rohde-Hildreth device, the questions to be answered are: Does the test serve the function it was intended to serve? Is it useful or helpful in any way in studying or appraising the individual? It it a reliable instrument?

The writer has done the following described experimental work for test validation. The subjects chosen for the validation study were ninth grade high school students in public schools, including altogether 670 students from several high school populations. This grade was chosen because at this age level students are in a critical stage of personal adjustment, and understanding of the adolescent's personality, problems, conflicts, and aspirations is a highly important function of teachers and counselors. The average age of these pupils was fifteen years.

Murray's conceptual scheme of personality (7) provided the essential concepts of dynamic behavior units and variables for interpretation and evaluation of experimental data obtained with the completion test. All test blanks were analyzed in terms of the Murray categories.

The questionnaire responses were referred to a list of 39 personality categories based on the Murray schema, including such items as abasement, aggression, creativity, deference, dominance, flight from danger or harmavoidance, passivity, rejection, seclusion, sex, conflict, emotionality, optimism, and Superego (conscience).

Using this classification and description of variables as a guide, it was possible to translate the sentence completion data into quantitative results. This was accomplished, considering the test as a whole, first by analyzing the contents of the sentence completions for each individual into needs, inner states, and press; and secondly by determining their frequencies and intensities. Frequency and intensity in a given subject's responses were estimated as follows: Frequency was obtained merely by counting the number of occurrences of each variable. In

estimating intensity, a one to three scale was used: 1. low, 2. average, 3. high, the criteria for judging being the vividness and potency with which a variable was expressed.

For purposes of scoring, it was deemed best to score only gross categories and the main subdivisions: the major forms of needs, inner states, and press.

From the following brief excerpts of the papers written by two ninth grade boys, designated "A" and "B," one may gain some idea of the method of evaluating and of scoring the responses. One may also obtain a fairly good notion of the personality of the boys, although less than one-third of their answers are quoted. The varying needs and press indicated by responses to the same sentence beginnings are shown since similar numbers were extracted from each paper.

"A's" paper was written by an orphan boy who had lost both parents. He lived in a children's home in New York City and was well thought of by his teachers and school mates. He studied hard, was much interested in art, but disliked algebra. His marks in art, English and social studies were above average, while he managed to get barely a passing grade in algebra. After graduation from the junior high school, he was admitted to the School for Music and Art, a senior high school, where he pursued his work with keen enjoyment. It is significant that his great need for adventure, excitance and change impelled him to enlist and join the armed forces three months before graduation when he learned that diplomas would be granted to the boys who had fulfilled the requirements up to that time. Thus, further follow-up was prevented.

Subject "A" Age: 15 yrs. 7 mos. I.Q. 114

1. I want to know *if all have the same feeling for art as I do.*
2. The future *seems very bright and cheerful.*
3. My school work *has been very interesting to me.*
4. Earning my living *is a thrill.*
5. My greatest longing *is to paint.*
6. Secretly *I steal food from the pantry.*
7. If I *fail in algebra, I would practically "die."*
8. There are times *when I feel like running away and start a new life.*
9. Work is *pleasant and hard.*
10. Friends *are a help and encouragement.*
11. I become *embarrassed when I can't dance.*
12. Girls *fascinate me.*
13. Love *is grand!*
14. Other people *are quite interesting.*
15. The laws we have *are sometimes unjust.*
16. I cannot understand what makes me *so nervous and stammer.*
17. My stomach *is fine and holds a lot.*
18. At night *I study.*
19. My mother *is dead!*
20. Death *is sometimes inviting.*

An analysis [4] of the above responses gives one the impression of a worried, mentally stimulated, but yet relatively normal and adjusted boy who has held his own despite some severe handicaps. He appears to be slightly below average in scholarship. His preference for art as compared with algebra marks him as one with a verbal or pictorial mind rather than mathematical intelligence, and one who would not get universally good grades. The fact that his attitude as expressed here is amiable and optimistic, despite his obstacles, suggests a fairly well-structured ego, built up by pride and reinforced by a satisfactory conscience.

The outstanding press (environmental forces) indicated are:

p Affiliation—Sentence Nos. 10, 12, 14.
p Aggression (Belittlement or ridicule, possibly on account of his stuttering, inability to dance, and failures in algebra, plus some inferiority in regard to his speech)—Sen. Nos. 7, 11 and 16.
p Dominance—Sen. No. 15.
p Imposed task—Sen. Nos. 7, 9, 18.
p Loss of support (Death of mother and perhaps lack of nurturance from sub-parents)—Sen. No. 19.

The most important needs (n) and traits expressed are:

n Abasement (Masochism, thoughts of suicide, and leaving the field of failure)—Sen. Nos. 7, 8 and 20.
n Achievement (Interest in schoolwork, earning a living, and painting) fused with a good deal of n Counteraction (Desire to overcome weaknesses and obstacles)—Sen. Nos. 3, 4, 5, 9, 18.
n Acquisition (Desire to earn money, and stealing food out of lack)—Sen. Nos. 4, 6.
n Autonomy and freedom (Motility, change, excitance) fused with n Inviolacy (Intolerance of failure—i.e., leaving field of failure), also with n Cognizance, n Abasement and n Harmavoidance (low)—Sen. Nos. 8 and 15.
n Affiliation fused with some n Succorance—Sen. Nos. 10, 12, 14.
n Cognizance (Curiosity and desire for knowledge, interest in schoolwork, travel, and new scenes) fused with n Autonomy and freedom—Sen. Nos. 1, 3, 8 and 14.
n Harmavoidance (Low, repressed by pride)—Sen. Nos. 7 and 8.
n Inviolacy (Pride and intolerance of belittlement and failure) fused with n Infavoidance (Desire to escape shame and humiliation)—Sen. Nos. 7, 8, 11 and 16.
n Nutrition—Sen. Nos. 6 and 17.
n Sentience (Interest in beauty, color and other senuous impressions)—Sen. Nos. 1 and 5.
n Sex—Sen. Nos. 12 and 13.

The outstanding trait and inner states distinguished are: Emotionality—Sen. Nos. 1, 2, 4, 8, 16. Optimism and cheerfulness—Sen. Nos. 2, 4, 9, 10 and 17.

[4] This analysis is a partial quotation from the blind analysis made for the writer by H. A. Murray who also gave other valuable suggestions in connection with this study.

The Ego Ideal is to become an artist—Sen. Nos. 1 and 5.

The positive cathexes (power of objects to arouse responses of a certain kind in subjects) expressed are: art, food, friends, girls, painting and sometimes death. The negative cathexis is algebra.

The personality profile given in part above was corroborated, as were those of other subjects studied, by interviews with the subject himself, with teachers, administrators, schoolmates, and others who had been associated with the boy.

In contrast with the frank, cheerful, optimistic person just discussed, we find in the following responses of subject "B" an attitude of dark pessimism, envy, distrust and antagonism, a tendency to make idealistic statements that have little support, and to do much wishful thinking rather than striving to achieve his ambitions as indicated by an analysis of the needs, inner states and press.

Subject "B" Age: 14 years 3 months I.Q. 118

1. I want to know *why mankind is destroying itself.*
2. The future *is an infinite darkness.*
3. My school work *has been improving of late.*
4. Earning my living *by science is my greatest ambition.*
5. My greatest longing *is to master science.*
6. Secretly *I hope I will improve physically.*
7. If I *can become an optometrist, my ambition will be realized.*
8. There are times *when I wish I already had a full knowledge of science.*
9. Work is *unfair under all present regimes.*
10. Friends *are a beauty of nature.*
11. I become embarrassed (*not answered*).
12. Girls *are better companions than boys.*
13. Love (*not answered*).
14. Other people *are interesting specimens of personality.*
15. The laws we have *are the best in the world but still not perfect.*
16. I cannot understand what makes me so *poor in spelling even if I study two hours straight.*
17. My stomach *don't agree with sardines.*
18. At night *I journey into the vast beyond in wild adventure.*
19. My mother *loves my brother more than me.*
20. Death *is a great adventure.*

The outstanding needs (n) and traits are:

n Achievement—Sentence Nos. 3 and 16—fused with n Counteraction (Desire to overcome obstacles)—is not sufficient to support his Ego Ideal (high aspirations)—Sen. Nos. 4, 5, 7 and 8.

n Affiliation is fused with n Fantasy—Sen. No. 10—and becomes empty sentimentality when considered with Sen. No. 12 which expresses a fusion of n Affiliation, n Aggression (Jealousy of possible rivals among boys) and n Sex. (Cf. also n Aggression.)

n Aggression (Envy of brother—Sen. No. 19—and antagonism towards social-economic conditions—Sen. No. 9—plus the Narcissistic attitude towards other people—Sen. No. 14) further points to lack of sincerity in friendship.

n Cognition (Desire for knowledge) is high—Sen. Nos. 1, 4, 5, 7 and 8—but is stagnated by n Fantasy (Desire to escape from reality). (Cf. n Fantasy and sentence references below.)
n Counteraction (Desire to overcome weakness) is concerned with his physical condition rather than moral weaknesses—Sen. No. 6.
n Fantasy and irreality (Desire to dream or to do autistic thinking)—Sentence Nos. 8, 18 and 20. The desire for perfectionism—Sen. No. 15—is a further manifestation of n Fantasy which here is fused with n Deference (Praise of laws).
n Seclusion and secrecy (Failure to answer Sen. Nos. 11 and 13) also suggest evasion and lack of sincerity.
Pessimism and dejection probably are a projection of bitterness resulting from antagonism towards his mother, jealousy of his brother and other boys, and self-centered attitude toward people and the world in general—Sen. Nos. 1 and 2.

The important press (p) are:

p Lack of support (Mother's preference for brother)—Sen. No. 19.
p Foreboding anxiety—Sen. Nos. 1 and 2.
p Imposed task—Sen. Nos. 3 and 16.
p Physical affliction—Sen. No. 6.

The positive cathexes are: science, friends, girls, and death.

The negative cathexes are: work, present regimes, boys, sardines, mother, brother and probably spelling and schoolwork. The latter it may be noted outweight the former.

Interviews with teachers and supervisors disclosed that the boy was considered brilliant but lazy, therefore made poor grades in most of his subjects with the exception of one or two things which he especially liked. In a personal interview he criticized all of his teachers, his mother and brother, and all social and economic conditions. There was no evidence of ill-health. On the contrary, he looked strong and healthy. Some tactful questioning brought forth the information that the great desire to build up physically stemmed from self-admiration and the wish to be attractive to the opposite sex. He was reported to be indulged by his parents. In consequence of lack of understanding and pampering, the boy had developed thinking habits which prevented satisfactory progress in school.

To translate the judgments of teachers, counselors, and the experimenter into quantitative terms, the rating for every variable for each student was determined on a scale of one to ten. The score for each variable was then found by averaging the independent ratings of the judges.

To determine the validity of these analyses and the interpretations, the scores obtained from random samples of test paper responses of fifty girls and fifty boys were correlated with the combined ratings obtained from teacher's judgments of the subjects, the experimenter's interview

Table 1

Showing Correlation Coefficients between Ratings of Student's Responses and Ratings of Combined Judgments of Teachers and Other Sources Previously Mentioned (Columns 1 and 2), and between Ratings of Initial and Final Tests (Columns 3 and 4—Corrected for Attenuation)

Needs	Girls	Boys	Girls	Boys
Abasement	.89	.89	.97	.70
Achievement	.49	.59	.85	.71
Acquisition	.87	.75	.55	.77
Affiliation	.65	.77	.91	.85
Aggression	.85	.85	.56	.62
Autonomy	.90	.90	.72	.70
Blame avoidance	.81	.83	.74	.63
Change	.96	.94	.87	.73
Cognition	.92	.92	.94	.95
Counteraction	.53	.54	.75	.47
Deference	.85	.76	.82	.82
Defendance	.93	.72	.85	.82
Dominance	.30	.53	.65	.82
Exhibition	.78	.74	.90	.93
Exposition	.60	.87	.91	.98
Fantasy	.94	.89	.79	.84
Infavoidance	.95	.86	.81	.64
Nurturance	.69	.88	.88	.84
Organization	.53	.85	.84	.97
Playmirth	.63	.86	.85	.74
Recognition	.86	.75	.84	.80
Seclusion	.85	.93	.97	.94
Sentience	.69	.91	.90	.67
Sex	.69	.80	.84	.44
Succorance	.88	.83	.50	.84
Inner States				
Conflict, Perplexity	.82	.91	.76	.39
Dejection, Pessimism	.93	.88	.85	.56
Exultation, Optimism	.93	.93	.92	.63
Emotionality (Trait)	.89	.86	.72	.81
Inner Integrates				
Ego Ideal Achievement	.90	.90	.58	.75
Ego Ideal Pride	.91	.82	.89	.78
Superego	.88	.96	.74	.80
Narcism	.72	.80	.83	.84
Averages*	.79	.82	.82	.76

* The difficulty in securing adequate validation of such variables as Nutriance, Passivity and Rejection by teacher ratings necessitated the omission of these variables in the table. These variables were included, however, in the computations to determine consistency of response, but are not indicated in the table shown above which is an abridgment of the four original tables. Averages given in columns 3 and 4 were taken from the original table.

with them and their parents, the opinions of school administrators and guidance counselors concerning the students, and from their school records.

These scores and the test paper scores were converted to equivalent scores for computation of correlation coefficients.

The linear correlations (Pearson product-moment) between ratings of students' responses in the sentence completions items and the ratings of the combined judgments of teachers, the experimenter's interview ratings and other sources previously mentioned, all categories combined, were as follows: 0.79 for the girls, and 0.82 for the boys. For the different categories the correlations varied from 0.95 to 0.295. The lowest correlation was found for the need dominance. The result suggests that the test may not provide adequate stimuli for the expression of this need. Table 1 shows *r*'s for variables.

To obtain an estimate of the reliability of scoring the sentence completion responses, the scorings of from one to four raters in addition to the original experimenter were obtained. All these judges were thoroughly familiar with Murray's conceptual scheme of variables. There was a 95.5 per cent agreement in the items scored by the experimenter and an independent rater on 36 papers, while there was a 78 per cent agreement when five judges rated the same items on twelve papers.

Consistency of responses for 21 girls and 23 boys was retested after eight months. The averages of obtained *r*'s between initial and final test scores, corrected for attenuation by the Spearman-Brown formula, were 0.82 for the girls, and 0.76 for the boys.

The greatest changes after eight months were found in the individuals who were not well adjusted emotionally and socially. Apparently the unadjusted person who has unresolved problems strives unceasingly, first in one way then another, to find solutions. The needs of the adjusted person appear to be more stable.

It is impossible to give here a full account of the experimental results, to indicate the typical problems, fears, wishes, and difficulties that characterize the ordinary ninth grade high school students.

The study revealed that teachers seemed to be capable in observing objective needs arising from environmental conditions. Deep-seated emotional conflicts were less readily understood. In some cases maladjusted students were considered queer, but the underlying causes were not suspected by parents or teachers.

Thirty per cent of the girls and 32 per cent of the boys expressed concern about passing in their studies. A large per cent had worries connected with school. The greatest longing expressed by these students was to attain their Ego Ideal. Doing something interesting or thrilling was the greatest longing of 20 per cent of both boys and girls.

Received March 20, 1945.

References

1. Frank, L. K. Projective methods for the study of personality. *J. Psychol.*, 1939, **8**, 389–413.
2. Frank, L. K. Projective methods for the study of personality. *Trans. N. Y. Acad. Sci.*, 1939, **1**, 129–132.
3. Horowitz, Ruth, and Murphy, Lois B. Projective techniques in the psychological study of children. *J. exp. Educ.*, 1938, **7**, 133–140.
4. Lorge, I., and Thorndike, E. L. The value of the response in a completions test as indications of personal traits. *J. appl. Psychol.*, 1941, **25**, 191–199.
5. Lorge, I., and Thorndike, E. L. The value of responses in a free associations test as indicators of personal traits. *J. appl. Psychol.*, 1941, **25**, 191–199; 200–201.
6. Murray, H. A. Techniques for a systematic investigation of fantasy. *J. Psychol.*, 1936, **3**, 115–143.
7. Murray, H. A., et al. *Explorations in personality.* New York: Oxford University Press, 1938.
8. Payne, A. F. *Sentence completions.* New York Guidance Clinic, 1928.
9. Sanford, R. N., et al. Physique, personality and scholarship. Washington, D. C. *Soc. Res. Child Devel., National Research Council*, 1943.
10. Symonds, P. M., and Samuel, E. A. Projective methods in the study of personality. *Rev. Educ. Res.*, 1941, **11**, 80–93.
11. Symonds, P. M., and Krugman, M. Projective methods in the study of personality. *Rev. Educ. Res.*, 1944, **14**, 81–98.
12. Tendler, A. D. A preliminary report on a test for emotional insight. *J. appl. Psychol.*, 1930, **14**, 123–136.

Speech Intelligibility Under Various Degrees of Anoxia *

G. M. Smith and C. P. Seitz †

College of the City of New York

This study is an extension and a refinement of an earlier study by the same investigators (3). In the earlier study it was demonstrated that an altitude of 18,500 ft., simulated in a nitrogen dilution chamber, produced a statistically reliable decrement in speech intelligibility under certain conditions of initial difficulty. It was pointed out that the magnitude of the decrement, though not large, was indicative of a change that, under trying conditions, could become hazardous in commercial and military aviation. Though it is common practice to use oxygen above 10,000 or 12,000 ft., a pilot flying at 40,000 to 45,000 ft. breathing pure oxygen has, because of the reduced oxygen tension in the lungs, a physiology which roughly corresponds to that of a man flying at 16,000 to 18,000 ft. without oxygen. The demonstration of a decrease in speech intelligibility in this general altitude range is, therefore, not without some practical significance.

It was felt that the effect of altitude on speech intelligibility observed in the previous study would have been greater had some more adequate means been found to equate the difficulty of the test materials used in the altitude runs with those used in the control runs. It also seemed desirable to investigate the possibility of a decrement in speech intelligibility at altitudes somewhat lower than 18,500 ft., the altitude selected in the first investigation. The purpose of the present study was to make such an investigation with improved techniques.

Method and Materials

The method of observing the effect of oxygen deprivation in the present study involved the subjects' ability to hear and to indicate on a check list words in common speech. The use of a check list by the subjects was an innovation, for in the earlier study the subjects reported to the experimenter who checked their responses for them. The new pro-

* The authors are indebted to the Linde Air Products Co. for a liberal grant of oxygen and nitrogen, and to Messrs. Mortimer Feinberg and Max Rosenbaum for valued clerical and statistical assistance.

† On leave, Lieutenant U.S.N.R. The opinions or assertions contained herein are the private ones of the writers and are not to be construed as official or reflecting the views of the Navy Department or the Naval Service at large.

cedure eliminated the possibility of errors being introduced by the experimenter. The possibility that errors on the subjects' check list might be errors in visual perception and in checking rather than errors in auditory perception was minimized by the use of a procedure described below.

Uniformity in the difficulty of the test materials used during the control and the three experimental sessions was assured by the use of eight lists of recorded stimulus words. Test batteries used for any one of the four conditions were made up of the same lists of recorded test words, but they were presented in varying order. The test series were made up in part from standard word lists developed by the Bell Telephone Laboratories (2), covering the more frequent sounds that occur in common speech. Other test series were constructed by the investigators on the principles employed by the Bell Laboratories. Intelligibility for vowel sounds was tested by the use of lists of monosyllables all having the same initial and final consonants in any one list; e.g., *suit, sit, sat, set*, etc. Consonant intelligibility was tested by similar lists, each involving a constant vowel sound but a variation in the initial or final consonant; e.g., *nor, bore, yore, more*, etc. In each list there were 11 vowel items and 24 consonant items.

The recordings of the test materials were made on high fidelity equipment at the National Broadcasting Co. Studios in New York City.[1] To minimize the effect of wear these recordings were later put in semi-permanent form by the RCA Manufacturing Co. of Camden, N. J.[2] The recordings were played back through a Fairchild pick-up [3] coupled with a Presto amplifier [4] (especially adapted so as to give a relatively flat response curve) and Western Electric earphones.[5]

The experiments were carried out in the nitrogen dilution chamber of the College of the City of New York, which has a capacity of approximately 450 cu. ft. and walls approximately 8 in. thick, which provided adequate sound-proofing. An air conditioning unit maintained constant and comfortable temperature and humidity conditions, 74° F. and 60% relative humidity. The three simulated altitudes were 13,600 ft., 16,900 ft., and 20,100 ft. (corresponding to oxygen percentages of 12.5, 10.3, and 8.85, respectively). Samples of the chamber air were taken after altitude had been attained in each run and were analyzed on the Haldane-Henderson-Bailey gas analysis apparatus. The samples did not deviate from

[1] We are indebted to the National Broadcasting Co. and to Mr. R. A. Lynn of the engineering department for their cooperation.

[2] We are indebted to the RCA Manufacturing Co. and to Mr. W. L. Tesch of the record engineering department for their courtesy.

[3] Turntable unit model 199 and pick-up model 209.

[4] Model 87B.

[5] Type 588A.

the desired altitudes by more than 600 ft., with one exception.[6] The CO_2 content of the chamber air averaged 0.52%.

Subjects

Twelve male college students free from systemic defects as determined by medical examination served as subjects. They were supervised during each experimental run by a physician [7] in addition to the experimenter. The ages ranged from 18 to 21 years, with the median at 19 years.

Procedure

The subjects worked in four groups of three, each group being tested under four separate conditions: sea level and three altitudes, 13,600 ft., 16,900 ft., and 20,100 ft. On the sea level run all the usual routine for altitude runs, including the use of an oxygen mask by the experimenter, was followed; so the subjects were not aware that it was a control run. The usual precautions to allay fear of the chamber were taken. During the testing subjects were comfortably seated and were equipped with standard Western Electric earphones such as are used by the American Airlines on transport planes.[8] Record booklets made up of twelve separate tests, each containing check rows for eleven vowels and twenty-four consonants, were employed.

Table 1

Test Order

Test Number	Sound Level	Condition
Instructions	High (30 db.)*	Sea Level
1	High	Sea Level
2	High	Altitude**
3–10 (main series)	Low (24 db.)***	Altitude**
11	High	Altitude**
12	High	Sea Level

* Vowel articulation was approximately 98% and consonant articulation was approximately 94% at sea level.

** On control runs this was a pretense only.

*** Vowel articulation was approximately 92% and consonant articulation was approximately 51% at sea level.

[6] During part of one of the four 16,900 ft. runs the altitude dropped to 15,800 ft.

[7] We are indebted to Dr. William Recht for medical supervision.

[8] Type 588A. The response curves of one set of phones were plotted at the Stevens Institute of Technology through the courtesy of President H. N. Davis and Dr. H. Burris-Meyer. Though the curves are peaked rather sharply at 1,000 cps., they are relatively flat (±10 db.) between 2,000 and 8,000 cps. The transmission system as a whole was sufficiently free from distortion to make possible a ready identification of the speakers' voices. We are indebted to the American Airlines and to Messrs. D. W. Rentzel and H. A. Wolfe for the use of the phone sets.

The order and condition of test administration is indicated in Table 1. The first test, following the recorded instructions, was given at sea level at a sound intensity level so high that it offered almost no difficulty in intelligibility (see first footnote under Table 1). The second test was administered at altitude [9] at the same high sound level, after a 15 minute period of adjustment. Then the main series of tests, tests 3-10, followed at a much lower sound level, intentionally set so that the consonant articulation (the percentage of correct responses) was in the neighborhood of 50% (see third footnote under Table 1). Following the main test battery, the sound level was again increased to the initial high value and test 11 was taken while still at altitude. Finally, following return to sea level, test 12 was administered at the same high sound level. The purpose of these initial and final tests at the high sound level was to provide a check on the possibility of errors due to the subjects' inability to manipulate pencils accurately or to perceive the response on the check list in the five seconds allowed for this purpose.

The experimenter remained in the chamber throughout the run to supervise the procedure, to check on the condition of the subjects, and to allay any apprehension. The average time of ascent was 14 minutes, the time for the administration of the tests was approximately 45 minutes; thus, the total time in the chamber, including the 15 minute period of adjustment, was approximately 75 minutes.

The test booklets were in four forms, one for each of the four sessions. Identity of difficulty with a minimal practice effect was achieved by varying the order in which the recordings were played back. The order in which the four different conditions in the chamber occurred was different for each of the four groups. This was arranged in such a way that any advantage attributable to order per se was reduced to a minimum.

Results

The principal data are summarized in Table 2 which gives the articulation values for vowels, consonants and standard syllables for each of the twelve subjects, at sea level and at the three altitudes, for the main series of tests (tests 3 to 10 combined), which were administered at the low sound level. Standard syllable articulation values were calculated by the Fletcher and Steinberg formula $S = 1-(1-VC^2)^{0.9}$ derived empirically from extensive observations in the Bell Laboratories. The means and the probable errors of these means are also presented in this table. As is well known, the articulation values for vowels are very much better than for consonants under all conditions. The mean values for vowels, consonants and standard syllables all drop systematically as the altitude

[9] The ascent was of course a pretense on the control runs.

Table 2

Articulation Values for Vowels, Consonants, and Standard Syllables Tests 3 to 10 Combined* (Low Sound Level)

Subject	Sea Level			13,600 Feet		
	Vowel	Consonant	Syllable**	Vowel	Consonant	Syllable**
1	93.0%	49.5%	21.0%	84.0%	47.0%	17.0%
2	100.0	63.5	37.0	96.5	49.0	21.0
3	70.5	28.5	5.0	35.0	21.5	1.5
4	97.5	63.5	36.0	85.0	42.0	13.5
5	94.0	46.0	18.0	90.0	44.5	16.0
6	99.0	63.0	36.0	96.5	70.5	44.5
7	79.5	34.0	8.5	86.5	49.0	19.0
8	96.5	68.5	42.0	67.0	52.0	16.5
9	94.0	42.0	15.0	100.0	69.5	44.5
10	97.5	64.0	37.0	100.0	57.5	30.5
11	82.0	31.5	7.5	88.5	40.0	12.5
12	95.5	60.5	32.0	100.0	49.5	22.5
Mean	91.6	51.2	24.6	85.7	49.3	21.6
P.E.$_M$	1.7	3.1	3.0	2.9	1.9	2.3

Subject	16,900 Feet			20,100 Feet		
	Vowel	Consonant	Syllable**	Vowel	Consonant	Syllable**
1	71.5%	37.5%	9.0%	35.0%	14.5%	0.5%
2	97.5	49.5	22.0	36.5	26.5	2.5
3	65.0	29.5	5.0	57.0	25.0	3.0
4	63.5	25.0	3.5	54.5	29.5	4.5
5	41.0	17.0	1.0	48.0	16.5	1.0
6	68.0	25.0	3.5	54.5	19.5	2.0
7	63.5	15.5	1.5	17.0	2.0	0.0
8	51.0	21.0	2.0	7.0	0.0	0.0
9	48.0	8.5	0.5	13.5	0.5	0.0
10	99.0	60.5	33.5	94.0	52.5	23.5
11	52.0	26.0	3.0	41.0	12.0	0.5
12	90.0	37.5	11.5	61.5	35.0	7.0
Mean	67.5	29.4	8.0	43.3	19.5	3.7
P.E.$_M$	3.6	2.7	1.8	4.5	2.9	1.0

* Total number of test items equals 278 for each subject (190 consonants, 88 vowels).

** $S = 1 - (1 - VC^2)^{0.5}$.

increases. In the case of syllable articulation the mean value at sea level is 24.6%; at 13,600 ft., 21.6%, a negligible drop. But at 16,900 ft. a sharp decrement has occurred, bringing the mean to 8.0%, and at 20,100 ft. the mean has dropped to 3.7%. The values for vowel and consonant articulation show a similar trend.

These relations are graphically portrayed in Figure 1, which makes it clear that somewhere between 13,600 ft. and 16,900 ft. there is a critical value for the particular order of difficulty of intelligibility used in this experiment. It should be emphasized that the character of this curve is very definitely a function of the initial difficulty at sea level. This statement is supported, not only by the earlier experiment, but by the supplementary tests in the present experiment. In the earlier study, where the initial difficulty was less, the mean syllable articulation at sea level (55.5%) dropped only 8% at 18,500 ft., as compared with decrements of 16.6% at 16,900 ft. and 20.9% at 20,100 ft. in this study (tests 3–10 combined), where the mean syllable articulation at sea level was 24.6%.[10] In the supplementary tests of this study (tests 1, 2, 11, and 12) administered at the high sound level, the decrement at altitude is also much less. This is indicated in the bottom row of Table 3, which compares the mean

Table 3

Comparison of Mean Syllable Articulation Values

Means Compared	Tests Involved	Sound Level	Difference between Means	*t* Values*	*P* (Probability that difference is due to chance)**
S.L.–13,600 ft.	3 to 10	Low	24.6 − 21.6 = 3.0%	0.76	.46
S.L.–16,900 ft.	3 to 10	Low	24.6 − 8.0 = 16.6%	4.5	.001
S.L.–20,100 ft.	3 to 10	Low	24.6 − 3.7 = 20.9%	5.7	.0002
13,600–16,900 ft.	3 to 10	Low	21.6 − 8.0 = 13.6%	3.1	.010
16,900–20,100 ft.	3 to 10	Low	8.0 − 3.7 = 4.3%	2.6	.025
20,100 ft. loud–20,100 ft. soft	2+11*** vs. 3 to 10	High–Low	57.9 − 3.7 = 54.2%	8.6	<.00001
S.L. loud–20,100 ft. loud	1+12*** vs. 2+11***	Both High	65.3 − 57.9 = 7.4%	1.1	.29

* These values were calculated by R. A. Fisher's technique (1) for testing the reliability of a difference between the means of small correlated samples.

** *P* was calculated by Fisher's method from Student's tables of *t*. The difference may be regarded as reliable when *P* is .01 or less.

*** These two tests combined involved only 22 vowel and 48 consonant test items, as compared with the combination of tests 3–10, which involved 88 vowel and 190 consonant items.

[10] The more marked decrements in this study are in part due to the more sensitive procedures employed.

syllable articulation for tests 1 and 12 taken at sea level with the mean for tests 2 and 11 taken at 20,100 ft. Here the decrement of only 7.4% is clearly due to the fact that the initial difficulty at sea level was relatively slight, as indicated by a mean syllable articulation value of 65.3% in contrast with the mean value of 24.6% in the main series.

Table 3 verifies the impression given in Figure 1. The difference in intelligibility between 13,600 ft. and sea level is not statistically reliable;

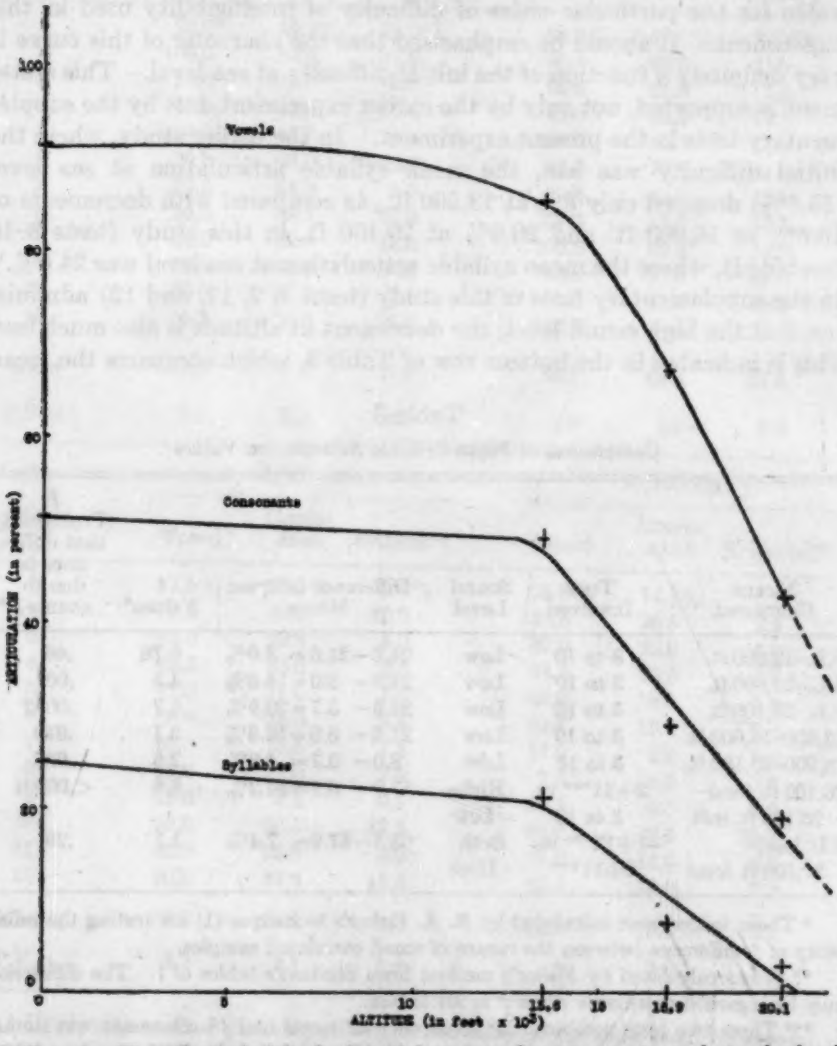


FIG. 1. The effect of altitude (anoxia) on vowels, consonants, and standard syllables. For the order of difficulty of intelligibility used in this experiment, a critical value appears between 13,600 ft. and 16,900 ft.

however, the differences in intelligibility between 16,900 ft. and sea level and between 20,100 ft. and sea level are very reliable. The difference between 16,900 ft. and 13,600 ft. is also reliable, and the difference between 20,100 ft. and 16,900 ft. is almost reliable. The degree of reliability of these differences is indicated by the column of P values, which indicate the probability that any obtained difference between means is due to chance. In general, the differences may be regarded as reliable for values of P equal to .01 or less. Even for $P = .02$ a fair reliability is indicated.

We get some light on the question of whether or not decreased articulation values at altitude are attributable to extra-auditory factors from a consideration of the next to the last row in Table 3. Here tests 2 and 11 presented at the high intensity level at 20,100 ft. are compared with the main series of tests (3–10) which were presented at the low sound intensity at the same altitude. The performance, as indicated by the highly reliable difference between the mean syllable articulation values, is vastly superior at the higher sound level. This indicates that the low articulation values obtained at 20,100 ft. in the main test series at the low sound intensity were not spurious; they were not merely the result of the subjects' inability to manipulate their pencils or to locate the proper items on the check lists, but were due to difficulties in auditory perception. That there was a small though not statistically reliable decrement in intelligibility at 20,100 ft. even at the high sound level is indicated in the last row of the same table. These results have already been discussed.

Individual differences in the effect of altitude on the ability to perceive speech sounds are shown in Table 4. The ratio of syllable articulation at sea level to syllable articulation at each of the three different altitudes is presented for each subject. A ratio greater than one indicates a decrease of intelligibility at altitude; a ratio less than one indicates an increase in intelligibility at altitude. It will be seen that even at 13,600 ft. 67% of the subjects show some loss of efficiency in perception, at 16,900 ft. 92% are affected, and at 20,100 ft. 100% are affected. Centering our attention on the ratios for the two higher altitudes, where the mean decrements have been shown to be statistically reliable, we observe a wide range in individual reactions to anoxia. In the third column the ratios vary from 1.00 to 30.0, indicating that syllable articulation for the most tolerant individual was the same at 16,900 ft. as at sea level, whereas the most susceptible individual had a syllable articulation at this altitude which was only 1/30 of what it was at sea level. In the last column of this table a similar wide range of ratios appears. In general, those subjects who were least affected at 16,900 ft. were least affected at 20,100 ft. and those who were most affected at the lower altitude were most affected at the higher altitude. As a check on the consistency of susceptibility

Table 4

Individual Differences in Reaction to Altitude: Ratios of Syllable Articulation Values at Sea Level to Values at Altitude

Subject	S.L./13,600 Ft.	S.L./16,900 Ft.	S.L./20,100 Ft.
1	1.2	2.3	42.0
2	1.8	1.7	14.8
3	3.3	1.0	1.7
4	2.7	10.3	8.0
5	1.1	18.0	18.0
6	0.8	10.3	18.0
7	0.5	5.7	—*
8	2.6	21.0	—*
9	0.3	30.0	—*
10	1.2	1.1	1.6
11	0.6	2.5	15.0
12	1.4	2.8	4.6
Mean	1.5	8.9	13.7*
Per Cent of Cases Showing Altitude Effect	67%	92%	100%

* Subjects 7, 8, and 9 obtained a syllable articulation value of 0 at 20,100 ft. The mean is based on the other 9 cases only.

to anoxia, as indicated by these ratios, the rank difference correlation between the values in the last two columns was calculated. It was found to be 0.70, with a probable error of 0.10.[11] This indicates that the rank order of the ratios S.L./16,900 ft. does not differ widely from that for the S.L./20,100 ft. ratios. This not only implies a fair degree of consistency in individual reactions to altitude, but it is a rough check on the reliability of the testing procedure.

Discussion

It should again be emphasized that the marked and reliable decrements in speech intelligibility observed at the two higher altitudes investigated in this study are not absolute values. They are definitely a function of the initial difficulty of the tests, as determined by the original speech production, by the fidelity of the recording and reproducing systems, and by the sound level employed. A question may be raised as to the significance of the results of the main series of tests which were intentionally run at a low sound level. The answer to this question is given in our earlier paper which reproduced a graph from the report of

[11] The rank difference correlation between the S values themselves for 16,900 ft. and 20,100 ft. is 0.77, with a P.E. of 0.08.

Fletcher and Steinberg (2) showing the relationship between syllable articulation and discrete word intelligibility. This indicates that when syllable articulation drops much below a point in the neighborhood of 30%, discrete word intelligibility falls precipitately and speech rapidly becomes unintelligible. It is further pointed out in our earlier report that no communication system permits perfect intelligibility under conditions of flight, because of cockpit noise, static, vibration, low pressure, poor speech production, and the special stresses of storm and combat. Hence, the additional factor of anoxia may reduce articulation, and consequently speech intelligibility, below the critical point.

Sixty-seven per cent of the subjects were affected at 13,600 ft. If such a decrement in speech intelligibility occurs in short flights of 1¼ hrs., one has additional evidence supporting the need for using oxygen on extended flights which, though generally not above 10–12,000 ft., may be of 10 to 12 hours duration. One hopes that at the higher altitudes oxygen equipment will be employed and employed effectively. With the practical introduction of pressurization the importance of anoxia as a problem for crews using conventional oxygen equipment and flying above 35,000 ft. should cease to exist.

Summary and Conclusions

Under standard conditions speech intelligibility is shown to decrease with increasing altitude. If the original sound level is low the decrement in articulation is large, discrete word intelligibility falls precipitately and speech rapidly becomes unintelligible. The effects of anoxia are minimal if sea level articulation is high. (As was pointed out in our earlier paper (3), this will be the case when: the sound level is high; a superior communication system is employed; personnel are trained in speech production and in listening under noisy and distracting conditions; and when "filling in" is made easy by familiarity with the type of message transmitted.)

The fact that 67% of the subjects (8 of 12) showed a decrement in performance at 13,600 ft. suggests the importance of using oxygen equipment on long flights even at low altitudes.

Received January 8, 1945.

References

1. Fisher, R. A. *Statistical methods for research workers.* London: Oliver and Boyd, 1938.
2. Fletcher, H., and Steinberg, J. C. *Articulation testing methods,* Bell Telephone Laboratories, Reprint B-436, November 1929, pp. 41, 42, 46.
3. Seitz, C. P., and Smith, G. M. Auditory sensitivity under conditions of anoxia; a study of speech intelligibility, *J. Aeronautical Sciences,* 1942, 9, 478–480.

Book Reviews

Peatman, J. G., and Hallonquist, Tore. *The patterning of listener attitudes toward radio broadcasts.* (Applied Psychology Monographs, No. 4.) Stanford University: Stanford University Press, 1945. Pp. 58. $1.00.

The principal purpose of this monograph is to illustrate the application of cluster analysis to Program Analyzer data. Each of two radio programs was divided into program units. Tetrachoric correlation coefficients (computed with the aid of Thurstone tables) were used to express the intercorrelation among units. Cluster analysis (after Tryon) was used to find whether the program units could be considered as members of one basic cluster or whether there were several factors in operation. The theme of the monograph appears to be that having a radio program with one basic audience mood (a one-factor show) is desirable and that this method provides a way of finding out whether a program does appeal mainly to one basic audience mood. The use of cluster analysis is proposed merely as a supplement to the usual treatment of Program Analyzer data.

Three estimates of reliability were made. Two of them were estimates of the reliability with which the show as a whole was measured. The third estimate was based upon the intercorrelation of program units within each cluster. The reliability coefficients for the clusters all were above .90 with most of them over .95.

The theme of the monograph, that a one-factor show is desirable from the standpoint of acceptance, is supported more by logic and experience than by the data presented. Results for only two programs were given. The one-cluster show had a much higher Program Analyzer rating than the five-cluster program. However, this result would hardly justify acceptance of the hypothesis.

Nevertheless, the hypothesis appears reasonable, and it is likely that the writers' proposal of this hypothesis was based on considerably more evidence than actually has been presented in this monograph. Anyway, no claim is made that having one unified pattern is the only important consideration.

The methodological questions raised directly or by implication are as interesting as the practical possibilities. These questions go beyond what might be called test-reliability. They would include: (1) the extent to which the responses in the Program Analyzer laboratory situation indicate the reactions of people listening in their homes—a problem which the

writers fully recognize; (2) the whole question of the extent to which a relatively small group of respondents, perhaps representing a limited geographic area and possibly subject to selective errors resulting from voluntary cooperation, can reflect the reactions of listeners in general; and (3) the accuracy of estimates of reliability made with the use of the Spearman-Brown formula for data of this type.

Some readers may feel that the statistical aspects have not been covered thoroughly. They may notice that there is no discussion of the assumptions underlying the formulae used, no raising of the question whether the statistical methods are appropriate for data of this type. They may expect, but fail to find, an evaluation of possible methods of factor analysis.

However, these questions could not have been covered thoroughly without deviating substantially from the principal theme. They are very important, but they are broader problems that probably should be covered separately.

The reviewer hopes that this monograph reports only the beginning of what should turn out to be a significant research program, including: extension of the method to additional problems, empirical tests of the hypothesis that a one-factor show is desirable from the standpoint of acceptance, and investigations of the statistical and methodological problems previously suggested. Such a research program surely is off to a good start.

Alfred C. Welch

Knox Reeves Advertising, Inc.,
Minneapolis, Minnesota

Inbau, Fred E. *Lie detection and criminal interrogation.* Baltimore: The Williams and Wilkins Company, 1942. Pp. viii + 142. $3.00.*

This book is intended primarily as a practical manual for criminal investigators. Its secondary objective is the stimulation of interest and research in the arts of lie detection and criminal interrogation.

The first section describes the technique of scientific lie detection using the Keeler Polygraph to detect the physiological concomitants of emotional reactions accompanying deception. The history, legal status, validity, and practical methodology of the technique are discussed. The effects of nervousness and fear, unresponsiveness, physiological and psychological abnormalities, and the like receive detailed treatment. Ingenious methods for meeting such contingencies are described.

According to Inbau, an accurate deception diagnosis can be made by a competent examiner in about 70% of the cases; while the indications

* The views expressed herein are the personal views of the writer and do not reflect the views of the United States Navy or the naval service in general.

may be too ambiguous or indefinite in about 20% of the cases to justify a definite statement. *And in about 10% of the cases the best examiner may be wrong.* Inbau takes this latter figure to represent the actual margin of error of the technique. He states that the chief source of error is in the failure to detect deception rather than in the misinterpretation of innocent truthfulness.

The second section describes practical techniques for criminal interrogation without lie detector aid. Suspects are classified according to certainty of guilt and probable degree of emotional conflict regarding the offense. Strategems appropriate to each class are presented.

As a practical manual this book is outstanding, but the lack of theoretical emphasis and sophistication may impair its research stimulating value. Inbau states, "The lie detector technique is predicated upon the theory that deception criteria appear in a record because of the emotional disturbances resulting from the subject's consciousness of lying and his fear of detection." But he also implicitly recognizes the importance of emotional conditioning or association, for instance, in his discussion of the possible after-effects of "third degree" examinations on the records of innocent, truthful persons. After such an experience, a truthful denial of guilt may be accompanied by emotional responses and polygraph records like those characteristic of deception. Moreover, Inbau states that persons who have successfully rationalized their crimes or who have been relieved of their emotional conflicts regarding an offense by confession to a clergyman may show few or no deception indices. And these indices of deception are absent while the criminal consciously lies concerning his offense presumably because of his fear of detection. Though "consciousness of lying" and "fear of detection" are important sources of emotional reactions, they may have been overemphasized in previous theoretical discussions of this subject. They seem to lack sufficient generality to be entirely suitable bases for either explicit or implicit theories of lie detection. Perhaps a theory formulated in terms of emotional conditioning or association might have both greater generality and greater stimulus value for psychologically meaningful experimentation.

Howard F. Hunt

United States Naval Reserve

De Silva, Harry R. *Why do we have automobile accidents?* New York: John Wiley & Sons, Inc., 1942. Pp. ix + 394. $4.00.

This book deals with some phases of the important problem of highway mishaps with emphasis on the psychological aspects of accident causation. An analysis is made of the contributing factors, such as exposure to accidents, speed, skill at the wheel and attitudes toward

safety. Data are given on the types of drivers behind the wheel and exposure-in-travel hazards. Characteristics of the driver, such as age, reaction time, experience and skill in general, are treated in relation to speed. It is stated that fifty miles an hour on the best of highways is fast enough for the average driver.

The author presents his own adaptation of tests for drivers which includes both road and laboratory procedures. Considerable attention is given to vision in relation to safe driving. The treatment in this section is somewhat stinted and overlooks the work done by many other investigators.

One section of the book is devoted to sociological aspects of accident susceptibility. Skill alone is said to be only a part of safe driving. Some evidence is given which indicates a higher accident ratio in the lower socio-economic groups. The author contends that illiteracy and foreign birth are added susceptibilities to accident, but no statistical analysis is offered to show that these factors operate independently of other conditions and attitudes. Consideration is given of the effects of intoxication, fatigue, ill health and related factors but the treatment is limited in scope.

A very satisfactory survey is made of methods used in collecting accident information, types of accidents, reporting systems, interpretation of data on accidents and criteria used in safety contests. Financial responsibility laws and compulsory insurance are discussed in a general way.

The last part of the book deals with various related problems, including pedestrian traffic, comparative figures on urban and rural accidents, enforcement, outstanding characteristics of drivers, factors concerning road conditions in relation to accidents and re-examination procedures of drivers. A research program on highway safety is presented at the conclusion of the book.

In all there are 374 pages and a bibliography of 300 listed references. A reasonable amount of tabular, graphic, and illustrative material is included. The index is rather incomplete, possible due in part to the wide coverage of topics within the volume. Although it might be much better organized, the book serves as an introduction to this very much neglected area of study, on the part of academic writers, and should be read by everyone interested in traffic problems and automobile accident prevention. It should, however, be read somewhat critically as the presentation is sketchy in places. The author describes his treatment in the preface as an attempt, "to offer the reader a bird's-eye view of the field."

A. R. Lauer

Iowa State College,
Ames, Iowa

New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota

Job safety training manual. Kenneth L. Faist, National Foremen's Institute, Inc., Deep River, Conn., 1944. Pp. 52. $5.00.

Training for supervision in industry. George H. Fern. New York: McGraw-Hill Book Co., Inc., 1945. Pp. 188. $2.00.

Understanding labor. Bernard H. Fitzpatrick. New York: McGraw-Hill Book Co., Inc., 1945. Pp. 179. $2.00.

Conference leadership in business and industry. Earle S. Hannaford. New York: McGraw-Hill Book Co., Inc., 1945. Pp. 289. $3.00.

Top-management planning. Edward H. Hempel. New York: Harper & Brothers, 1945. Pp. 414. $4.50.

How to get the job you want! John W. Herdegen. New York: Essential Books, Inc., 1945. Pp. 92. $1.00.

Sell America into jobs. William E. Holler. Motor City Publishing Co., General Motors Bldg., Detroit 2, Mich., 1945. Pp. 128. $1.50.

New goals for old age. George Lawton. New York: Columbia University Press, 1943. Pp. 210. $2.75.

Human leadership in industry: The challenge of tomorrow. Sam A. Lewisohn. New York: Harper & Brothers, 1945. Pp. 112. $2.00.

The social problems of an industrial civilization. Elton Mayo. Division of Research, Harvard Business School, Soldiers Field, Boston 63, Mass., 1945. Pp. 150. $2.50.

Counseling service for industrial workers. Mary Palevsky. New York: Family Welfare Association of America, 1945. Pp. 51. $.60.

Diagnostic psychological testing, Vols. 1 & 2. David Rapaport. Chicago: The Year Book Publishers, Inc., 1945. Pp. 573. $6.50 per vol.

Counseling with returned servicemen. C. R. Rogers and J. L. Wallen. New York: McGraw-Hill Book Co., Inc., 1945. Pp. 165. $1.60.

Sex and the social order. G. H. Seward. New York: McGraw-Hill Book Co., Inc., 1945. Pp. 286. $3.50.

Where do people take their troubles? Lee R. Steiner. Boston: Houghton Mifflin Co., 1945. Pp. 263. $3.00.

Productive thinking. Max Wertheimer. New York: Harper & Bros., 1946. Pp. 224. $3.00.

The relation between illumination and visual efficiency—the effect of brightness contrast. H. C. Weston, Report No. 87, Industrial Health Research Board, Medical Research Council, London, England, 1945. H. M. Stationery Office, York House, Kingsway, London, W. C. 2. Pp. 35. 9d.

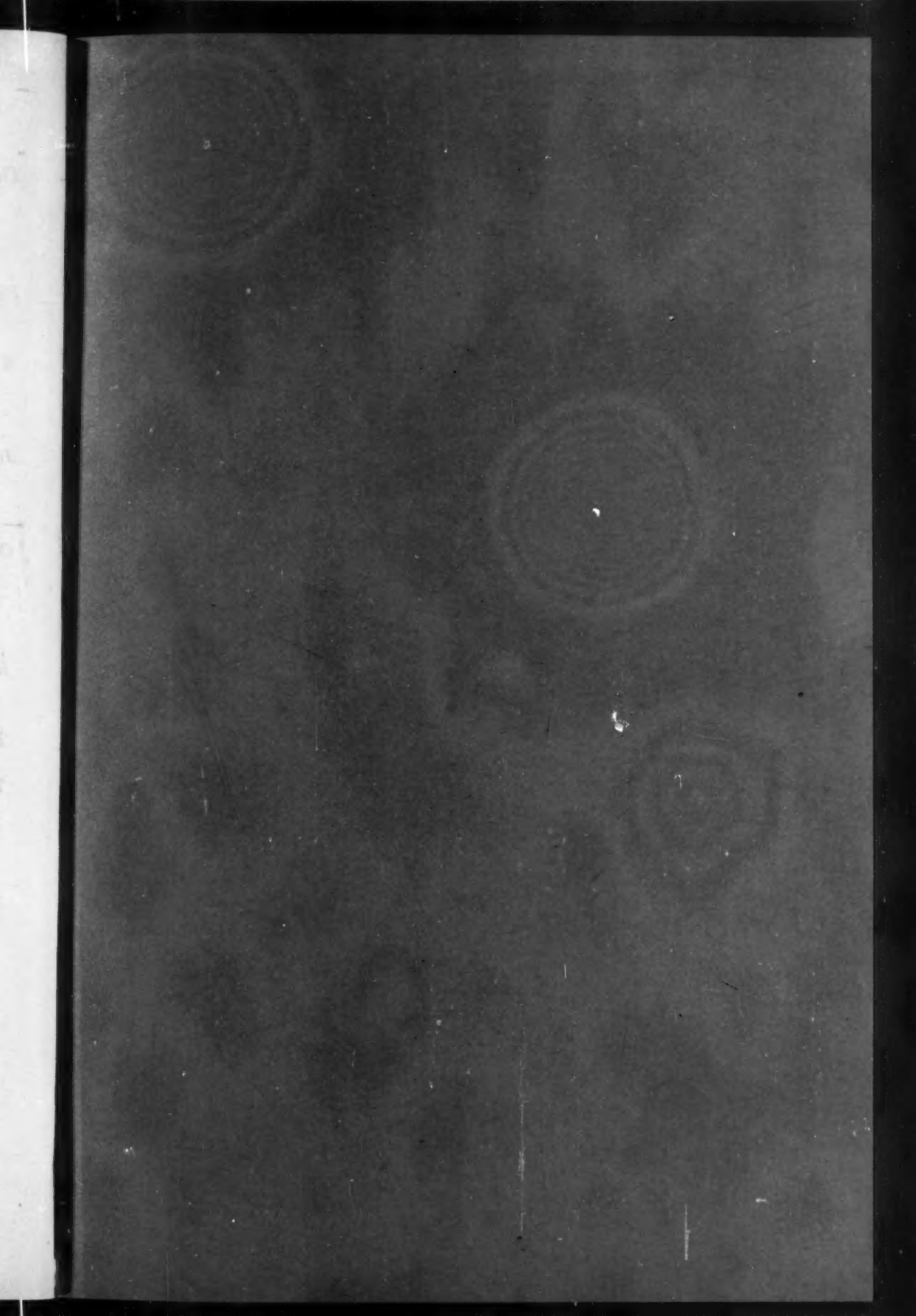
Procedures in evaluating a guidance program. Frances Morgan Wilson. New York: Bureau of Publications, Teachers College, Columbia University, 1945. Pp. 210. $2.60.

A study of women on war work in four factories. Industrial Health Research Board Report No. 88 of the Medical Research Council. S. Wyatt *et al.* H. M. Stationery Office, York House, Kingsway, London, W. C. 2, England, 1945. Pp. 44. 9d.

Job descriptions for office occupations. Washington 25, D. C.: Superintendent of Documents, U. S. Government Printing Office. 1945. Pp. 204. $1.25. Division of Occupational Analysis, War Manpower Commission.

Occupational data for counselors: A handbook of census information selected for use in guidance. Bulletin No. 817, Bureau of Labor Statistics, U. S. Department of Labor, 1945. Washington 25, D. C.: Superintendent of Documents, U. S. Government Printing Office, Pp. 36. $.10.

Racial aspects of reconversion: A memorandum prepared for the President of the United States. New York: National Urban League, 1945. Pp. 29. Gratis. (Cost of postage charged for bulk orders.)

The placement interview. Civilian Personnel Pamphlet No. 15. War Department, Washington 25, D. C. Pp. 17.

Training and research in industrial relations. Proceedings of May 1945 Conference. Bulletin 1. Industrial Relations Center, University of Minnesota. October, 1945. Pp. 1–60. Gratis.